

GenCore version 5.1.4 p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 16, 2003, 17:32:29 ; Search time 2217 seconds

(with alignments)
7487.765 Million cell updates/sec

Title: US-09-515-363c-2

Perfect score: 5311

Sequence: 1 MSGXSTDENFRYLISCFA.....LPTEPNDSYSCCLFSD 1025

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 16154066 segs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL-frame_plus_p2n.model -DEV-xlh
-Q/cgn2_1/USPTO/US09515363/runat_16052003_100247_248/'/app/que/y.fasta.1.1223
-DB-EST-QFMT-fastap -SUFIX-rst -MINMATCH-0.1 -LOCPL-0 -L-PEXT-0
-UNITS-bits -START-1 -END-1 -MATRIX-blosum62 -TRANS-human4.cdi -LIST-45
-DOCALLIGN-200 -NORM-SCORE-PCT -THR_MAX-100 -THR_MIN-0 -ALIGN-5 -MODE-LOCAL
-OUTFMT-oto -NORM-ext -HEAPSIZE-500 -MINLEN-0 -MAXLEN-200000000
-USER-US09515363-ecgn_1_1_1650-etunat_16052003_100247_24835 -NCPU-6 -ICPU-3
-NO_XLPLY -NO_MAP -LARGEQUERY -NEG_SCORES-0 -WAIT -LONGLOG DEV.TIMEOUT-120
-WARN_TIMEOUT-30 -THREADS-1 -XGAPOP-10 -XGAPEXT-0.5 -FGAPOP-5 -FGAPEXT-7
-YGAPOP-10 -YGAPEXT-0.5 -DELOP-6 -DELEXT-7

Database :
EST:
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rnd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Descr: job
1	1597	30.1	1013	14 B0233683	B0233683 AGENCOURT
2	1376	25.9	870	14 B0960157	B0960157 AGENCOURT
3	1218	22.9	1115	13 BM467983	BM467983 AGENCOURT
4	1206.5	22.7	1035	12 BF337464	BF337464 AGENCOURT
5	1129	21.3	729	12 BF983236	BF983236 AGENCOURT
6	1122	21.1	1197	11 AK018602	AK018602 Mus. muscu
7	1098	20.7	827	12 BG741146	BG741146 AGENCOURT
8	1094.5	20.6	781	12 BF686405	BF686405 AGENCOURT
9	1081.5	20.4	1041	12 BE882040	BE882040 AGENCOURT
10	1066	20.1	1046	13 BM476961	BM476961 AGENCOURT
11	1059	19.9	755	14 B0772836	B0772836 AGENCOURT
12	1055.5	19.9	1239	13 BM467774	BM467774 AGENCOURT
13	1016	19.1	672	14 B0316075	B0316075 AGENCOURT
14	1016	19.1	672	14 B0316108	B0316108 AGENCOURT
15	1016	19.1	672	14 B0316120	B0316120 AGENCOURT
16	990.5	18.6	911	13 B1454996	B1454996 AGENCOURT
17	989.5	18.6	1611	13 BM455238	BM455238 AGENCOURT
18	952.5	17.9	932	13 BG967530	BG967530 AGENCOURT
19	911.5	17.2	568	14 B0311714	B0311714 AGENCOURT
20	900	16.9	2398	11 BC027369	BC027369 Mus. muscu
21	896.5	16.9	790	12 BF160649	BF160649 AGENCOURT
22	892	16.8	1201	13 BM551088	BM551088 AGENCOURT
23	857.5	16.1	666	10 BB148072	BB148072 AGENCOURT
24	855	16.1	787	13 B1453288	B1453288 AGENCOURT
25	841.5	15.8	787	13 B1453288	B1453288 AGENCOURT
26	836.5	15.8	596	12 BF146945	BF146945 AGENCOURT
27	830.5	15.6	715	10 BB627907	BB627907 AGENCOURT
28	827.5	15.6	715	13 B1689931	B1689931 AGENCOURT
29	813.5	15.3	585	9 AA123960	AA123960 mp78b04.f
30	796	15.0	546	12 BE940626	BE940626 AGENCOURT
31	779	14.7	493	12 BG391418	BG391418 AGENCOURT
32	774.5	14.6	481	10 AM580584	AM580584 AGENCOURT
33	765	14.4	1319	14 BM806083	BM806083 AGENCOURT
34	751	14.1	536	9 AT171827	AT171827 as51e06.x
35	739.5	13.9	581	13 BM426861	BM426861 p452n.pk0
36	724	13.6	537	9 A1806204	A1806204 w126c12.x
37	722	13.6	521	9 A1922705	A1922705 w011f08.x
38	715	13.5	446	10 BE090624	BE090624 RC6-BT071
39	705	13.3	452	10 BE090630	BE090630 RC6-BT071
40	704	13.3	445	10 BE090627	BE090627 RC6-BT071
41	701	13.2	457	10 BE090628	BE090628 RC6-BT071
42	693	13.0	552	9 AA134958	AA134958 z023f04.s
43	672.5	12.7	520	10 BB284625	BB284625 AGENCOURT
44	670	12.6	538	10 AM189584	AM189584 x10bn12.x
45	660	12.4	484	10 AM152541	AM152541 x176d02.x

ALIGNMENTS

RESULT 1
B0233683
LOCUS B0233683
DEFINITION B0233683 1013 bp mRNA linear EST 02-MAY-2002
5' mRNA sequence.
ACCESSION B0233683
VERSION B0233683.1 GI:20415083
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euarchontia;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 1013)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.

Email: c94b05.r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Corporation (LIML)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LIML at:
http://image.liml.gov
plate: L1AM12875 row: g column: 07
High quality sequence stop: 721.

location/Qualifiers
1. .1013

BASE COUNT	ORIGIN
383	Average insert size 2.1 kb.
a	Site-2: SalI: cloned unidirectionally. Primer: Oligo dT
161	c
229	g
240	t
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone IMAGE:5786526"
	/clone_1lb="NH_MGC_71"
	/tissue_type="telomysarcoma"
	/lab_host="DH10B (phage-resistant)"
	/note="Organ: uterus; Vector: pCMV-SpOx56; Site:1: NotI;
	Site:2: SalI: cloned unidirectionally. Primer: Oligo dT

2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841

Pred. No.	1013
Length	322
Score	1597.00
Percent Similarity	95.61%
Best Local Similarity	94.15%
Query Match	30.07%
DB	14
Gaps	3

US-09-515-363C-2 (1-1025) x BQ233683 (1-1013)

QY	529	Lysansglnileglniuprocylslyslsphenalallelaspatlarratrgluasp	548
Db	2	AAAAACCAATACAGAGAGCCATCGAAGAAATTGGCTTCCATTCACATGCAATACAGAAACAT	61
QY	549	Prophelysgclulsleuleuglnilemetthratgilelnthtyrcysimelstcr	568
Db	62	CCATTTAAAGAAACCTTCTAGAAATATATGACAAGCATAAACCTATCTGCAATGACT	121
QY	569	Prometterasphegylthrglniprotyrgluglnintrpvalileglnmctglulslys	588
Db	122	CCAAATGTCACATTTTGGACATCAACCTATGAAACAATGGCCCATTCMAATGGCAAAAAA	181
QY	589	Alaialslslysglysnatrglyslarivaicysalacilunhsleuatclustyrasn	608
Db	182	GCTGCAAAACAGAAATTCGCAAGAACGCTGTTGTGCAAAACATTTGACCAAGTACAT	241
QY	609	Glnalaleuglnileasnaspthrileargmetlleaspatlarythrhlslsleuglntr	628
Db	242	GAGGCCCTACAAATATATGACACATTCGAATGATACATACGATATCTGCTTGAACAT	301
QY	629	Pheyltrsncluglulsasplyslsphenalavallelunhspspsaspgluclyl	648
Db	302	TTCATATATGAAAGAAAGATTAAGATTGGCTGATACACATGATACATGATGACGCT	361
QY	649	Glyaspaspglutryrcysaspsglyaspgluasgluaspspleuylsyrproleuyls	668
Db	362	GGTATATATGATTTGTGTGATGGTATGAAAGATGACGATATTTAAAGAAACCTTTGAAA	421
QY	669	Leuaspclunthraspatrgphenleumethleuphephegluasnsnlysmelleuyls	688
Db	422	CTGATGAAACAGATACATTTCTCATGACTTATTTTTTAAACCAATAAATAGTTGAAA	481
QY	689	Argleuualgluasnproglutryrgluasnglululsleuthrlyslsleuatqasnthrile	708
Db	482	AGGCTGGCTGAAAACCCAGAAATATGAAAATGAAAAGCTGACCAATTAATACCATTA	541
QY	709	Metclunltyrthrargthrnglunlseralargglylellephethrlysthrarg	728

Dd	542	ATGAGCAATAATACCTGGACACTGGAGCAATCAGCACGAGAATAATACCTTTCACAATAAAATACCA
Oy	729	GInserIatYrAlaleuSerGIInTrPIleThrgluAAnGIuLySPhaeIAcIuVAIGly
Dd	602	CAGAGTCATATCGCGCTTTCCAGTGATTTACTGAATAATGAAAAATTGGCTCAAGTAGGA
Oy	749	ValluysAlahIshIsLeuIIegIylagIyhISerSergIuPheLysPromoIThGln
Dd	662	GTCAAACCCCACCATCTGATGTGAGCTGCACACACAGCTGATTCAAAACCCATCAACAG
Oy	769	AasnGIuInuYSgluValIIeserLySPheaqThrgILySIleasnIeudeuIIAla
Dd	722	AATGAACAAAACAGACGATTAATGTAATTGCCACTGCAGAAAAATTAATCTGCTTAATGCT
Oy	789	ThrThryAlaIacIugluGlueasPIleLYsgIucYsaSnIIealIIeaqTYlly
Dd	782	ACCACAGTCGCACAAGAAAGCTCGCATATTAAAGAAATATACATGTGATTATCCGTAAAGT
Oy	809	LeuvalIhTrsnGIuIleaImetVAlgInaIaATgYAlaArGaIaAsPGIser
Dd	842	CTCGTCAC - AATGAATAAGCATGATGTCGAGGCCGTGGTCGAGGCACAGCTGATCAAGC
Oy	829	ThryrYalIeuvAlaIahISserGIySergIyValIIegIuhISgluThr----- Val
Dd	901	ACCTACCTCTCGTGTCTCTACGCGGCTTCA-----GAGTATCGAACCTAGCATATT
Oy	847	ASnaSPheaRgIuLySmelmeTyrlYsAlaIleHIScySaValgInasmoTIlsPro
Dd	952	AATCATTTCCGACA - AAGCATGATATATAG---CTATCATGGGGTTCAAAATGACAAACG
Oy	867	GIUGlu 868
Dd	1008	GAGGAG 1013

Source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6472264"
/clone_id="NH MGC 71"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: uterus; Vector: pCMV-SPORT6; size: 1.1 kb;
Site: 2; Salt; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2.1 kb."

```

BASE COUNT 319 a 131 c 204 g 214 t 2 others
 ORIGIN

Alignment Scores:
 Pred. No.: 2,67e-143 Length: 870
 Score: 1376.00 Matches: 279
 Percent Similarity: 97.90% Conservative: 1
 Best Local Similarity: 97.55% Mismatches: 2
 Query Match: 25.91% Indels: 5
 DB: 14 Gaps: 0

US-09-515-363c-2 (1-1025) x BM67983 (1-870)

592 LysGluAsnArgLysGluArgValCysAlaGluHisLeuArgLysTyrGlnGluValAla 611
 3 GAAGGAATTCGCAAAAGACCTGTTTGGCAAAAT-TTGGGAATACATGAGGCTGTA 61
 612 GlnIleAsnAspThrIleArgMetIleAspAlaTyrCHisGlnGluThrPheTyrAsu 631
 62 CAATTAATGACACAATTCGATGATGATGATGATGATGATGATGATGATGATGATGAT 121
 632 GluGluLysAspLysLysPheAlaValIleGluAspSerAspGluGluGluAsp 651
 122 GAAGAGAAAGATGAAGAGTTGAGCTCATGAGAGATGATGATGATGATGATGATGAT 181
 652 GluTyrCysAspGluAspGluAspGluAspAspLeuLysPheLeuLysLeuAspGlu 671
 182 GATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 241
 672 ThrAspArgPheLeuMetThrLeuPheGluAsnAsnLysMetLeuLysArgLeuAla 691
 242 ACAGATAGATTTCTCATGATTTATTTTGAACAATTAATGATGATGATGATGATGAT 301
 692 GluAsnProGluTyrGluAsnGluLysLeuThrLysLeuArgAsnThrIleMetGluGln 711
 302 GAAAGCCCAAGATATGAAGTGAAGAGCTGACCAATTAAGAAATACCAATGATGAGCAA 361
 712 TyrThrArgThrGluGluSerAlaArgGlyIleIlePheThrIleArgGlnSerAla 731
 362 TATACCTGAGCTGAGGATGAGGATGAGGATGATGATGATGATGATGATGATGATGAT 421
 732 TyrAlaLeuSerGlnTrpIleThrGluAsnGluLysPheAlaGluAlaLysAla 751
 422 TATGCGCTTCCAGTGGATTTACTGAAATGAAATTTCTGACAGTACAGTCAAAATC 481
 752 HisHisLeuIleGlyAlaGlyHisSerSerGluPheLysProMetThrGluAsnGluGln 771
 482 CACCATCTGATGAGCTGAGGACAGACAGTCAACCCAGACACAGATGAAACAA 541
 772 LysGluValIleSerLysPheArgThrGlyLysIleAsnLeuLeuAlaThrThrVal 791
 542 AAAGAGCTATTGTAATTTCCACCTGCAAAATTAATTTGATTCGATACCAACCTG 601
 792 AlaGluGluLysLeuAspIleLysGluCysAsnIleValIleArgTyrGlyLeuValThr 811
 602 GCAGAGAAAGCTCGATGATTAAGAATGTAACATTTGTTGCTGATGCTGCTGCTGCTG 661
 812 AsnGluIleAlaMetValGlnAlaArgGlyArgAlaArgAlaAspGluSerThrIleVal 831
 662 AATGAATATACCATGTCGCGCCGCGGCTGAGCCAGACATGATGATGATGATGATGAT 721
 832 LeuValAlaHisSerGlySerGlyValIleGluHisGluThrValAsnAspPheArgGlu 851
 722 CTGGTGGCTCAGAGTGGTTCAGAGTTCGAACTGACAGCAGTTAAATATTTCGACA 781
 851 LysMetMetTyrLysAlaIleHisCysValGlnAsnMetLysProGluGluTyrVal 870
 782 GAAGATGATGATTAAGCTATACATTTGTTCAAAATATGAACACAGAGATGATGATGAT 841
 870 HisLysIle 873
 842 TCATAGAT 851

RESULT 3
 BM67983 1115 bp mRNA linear EST 05-FEB-2002
 LOCUS
 DEFINITION AGENCOURT_6437921 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5532884
 5', mRNA sequence.
 ACCESSION
 VERSION BM67983.1 GI:18517025
 KEYWORDS
 SOURCE human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE
 1 (bases 1 to 1115)
 NIH-MGC <http://mhc.nci.nih.gov/>.
 TITLE
 JOURNAL
 COMMENT
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNI)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/ILNI at:
<http://image.llnl.gov>
 plate: L1AM12216 row: n column: 21
 High quality sequence stop: 690.
 Location/Qualifiers
 1..1115
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5532884"
 /clone_11b="NIH_MGC_71"
 /tissue_type="Telomysarcoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: uterus; Vector: pCMV-SPORE; Site: 1; Note:
 Site: 2; Salt: Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2.1 kb.
 BASE COUNT 326 a 221 c 218 g 277 t 73 others
 ORIGIN

Alignment Scores:
 Pred. No.: 2.19e-125 Length: 1115
 Score: 1218.00 Matches: 228
 Percent Similarity: 99.13% Conservative: 0
 Best Local Similarity: 99.13% Mismatches: 2
 Query Match: 22.93% Indels: 0
 DB: 13 Gaps: 0

US-09-515-363c-2 (1-1025) x BM67983 (1-1115)

796 LeuAspIleLysGluCysAsnIleValIleArgTyrGlyLeuValThrAsnGluValAla 815
 1 TTGGATATTAAGAAGATGTAACATTTGATTCCTTATGCTGCTGACCAATGAAATATC 60
 816 MetValGlnAlaArgGlyArgAlaArgAlaAspGluSerThrTyrValLeuValAlaHis 835
 61 ATGCTCCAGGCCCTGGTGTGAGCCAGAGCTGATGAGAGCACTACCTCTGTTCTTAC 120
 836 SerGlySerGlyValIleGluHisGluThrValAsnAspPheArgGluLysMetMetTyr 855
 121 AGGTGTCAGAGATTAATGACAGCTGAGACAGTATGATTTCCGAGAGAAAGATGAT 180
 856 LysAlaIleHisCysValGlnAsnMetLysProGluGluTyrAlaHisLysIleGluGlu 875
 181 AAAGATATCATTTGTTCAAAATATGAACACAGAGATATGCTCATATGATTTTTCGAA 240
 876 LeuGlnMetGlnSerIleMetGluLysLysMetLysThrLysArgAsnIleAlaLysHis 895
 241 TTACAGATGCAAGATATATGAAAGAAAGAAAGAAAGAAAGAAATATTTGTAAGCAT 300
 896 TyrLysAsnAsnProSerLeuIleThrPheLeuCysLysAsnCysSerValIleAlaLys 915

Db 301 TACAGATAAACCACATCACTTCTTCTTGCAGAACTGCAGTCTACTAGCTGT 360
 QY 916 SerGluGluAspIleHisValIleGluLysMetHisHisValAsnMetThrProGluPhe 935
 Db 361 TCTGGGAGATATCCATGTAAATTGAGAAATGCAATCATCATATGATTCAGAAATTC 420
 QY 936 LysGluLeuThrIleValArgGluAsnLysAlaLeuGlnLysLysGlyAlaAspTyrGln 955
 Db 421 AAGGACCTTACATGTGAGAGAAACAAACACTGCAGAAAGTGTGAGTACTGCA 480
 QY 956 IleAsnGlyLysIleLeuLysCysGlyGlnAlaTyrPheMetMetValHisLys 975
 Db 481 ATAAATGTCATATCATCTGCAATGTGGCAGGCTGGGAGACATGATGGTGACAAA 540
 QY 976 GlyLeuAspLeuProCysLeuLysIleArgAsnPheValValPheLysAsnAsnSer 995
 Db 541 GGCTTACATTTGCTGTCTGCAAAATAGCAATTTGTACTGCTTCAAAATATTC 600
 QY 996 ThrLysGlnIleLysIleLysIleValGluLeuProIleThrPheProAsnLeuAspTyr 1015
 Db 601 ACAAGAACATACAAAAAGTGGTAAATTAACCTATTCATTCATTCACAACTTGACTAT 660
 QY 1016 SerGluCysCysLeuPheSerAspGluAsp 1025
 Db 661 TCAGAAATGCTGTTATTACTGATGAGAT 690

RESULT 4
 BP337464 1035 bp mRNA line: EST 22-NOV-2000
 LOCUS 602035195F1 NCI_CGAP_Brn64 Homo sapiens CDNA clone IMAGE:4181126
 DEFINITION 5' mRNA sequence.
 ACCESSION BP337464
 VERSION BP337464.1 GI:11283715
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 NIH-MGC http://mgi.mcl.nih.gov/
 1 (bases 1 to 1035)
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgaab-remail.nih.gov
 Tissue Procurement: David N. Louis, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Place: LMAN9498 row: n column: 23
 High quality sequence start: 5
 High quality sequence stop: 695.
 Location/Qualifiers
 1. 1035

FEATURES
 source
 1. 1035
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4181126"
 /clone.lib="NCI_CGAP_Brn64"
 /tissue.type="glio"blastoma with EGFR amplification"
 /note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Clone unidirectionally. Primer: Oligo dT.
 Average insert size 1.57 kb. Constructed by Life
 Technologies. Note: this is a NCI_CGAP library."

BASE COUNT 353 a 221 c 243 g 218 t

Alignment Scores:
 Pred. No.: 3.78e-124 Length: 1035
 Score: 1206.50 Matches: 264
 Percent Similarity: 84.04% Conservative: 15

Best Local Similarity: 79.52% Mismatches: 44
 Query Match: 22.72% Indels: 12
 DB: 12 Gaps: 5
 US-09-515-363C-2 (1-1025) x BP337464 (1-1035)

QY 295 AsnValAlaAlaArgAlaSerProGluProGluLeuArgProTyrGln 414
 Db 6 AATGTGGCAGCAAGACATCCCGGAGCGAGAACTCCAGCTGAGCTTACTAAAGTAA 65
 QY 315 ValAlaGlnProAlaLeuGluGlyLysAsnIleIleLeuLysProThrLysGly 334
 Db 66 GTTGCCACCGACCGCTTGGAGAGAAATATCATCATCTGCTCCCTACAGGGAATGGA 125
 QY 335 LysThrArgValAlaValTyrIleAlaLysAspHisLeuAspLysLysLysLys 354
 Db 126 AAACCGAGATGCGTGTATACATGTCAGAGATCATAGCAAGCAAGAAAGACATCT 185
 QY 355 GluProGlyLysValIleValLeuValAsnLysValLeuLeuValGluLeuPheArg 374
 Db 186 GAGCTGGAAAGTATATGTTCTGTCAATAGGACTGCTAGTGAACAGCTCTTGGG 245
 QY 375 LysGluPheGlnProPheLeuLysTyrTyrArgValIleGlyLeuSerLysThr 394
 Db 246 AAGGACTTCCAAACCATTTTGAAGAAATGGTATCTGTTATTGATTAAGTGTATAC 305
 QY 395 GlnLeuLysIleSerPheProGluValValLysSerCysAspIleIleSerThrAla 414
 Db 306 CAACGTAAATATCATATTCACAGACTGTCAGAGTCTGTGATATATATCAAGT 365
 QY 415 GlnIleLeuGluAsnSerLeuLeuLysGluValGlnGlyValGlyLeu 434
 Db 366 CAATCTCTTGAAGACCTCTTAACTTGAAGTGAAGTGAAGAGATGCTGTAAATG 425
 QY 435 SerAspPheSerLeuIleIleLeuAspGlyCysHisIleThrAsnLysGluValArg 454
 Db 426 TCAGACTTTCCCTCATATATCATGATGATGATCATACACCAAGAAAGCAATGAT 485
 QY 455 AsnAsnIleMetArgHisTyrLeuMetGlnLysLeuLysAsnAsnArgLeuLysAla 474
 Db 486 AATACATCATCATGAGCATATATTTGATGCAAGATGCAAAACAAATAGACTCAAGAA 545
 QY 475 AsnLysProValIleProLeuProGlnIleLeuGlyLeuThrAlaSerProTyrGly 494
 Db 546 AACCAACCAAGTATTCCTCTTCAGATACGAGCACTTCACTGCTGCTTGGCA 605
 QY 495 GlyAlaThrLysGlnAlaLysAlaGluGlnHisIleLeuLysLeuCysAlaAsnLys 514
 Db 606 GGGCCACGACAGCAAGCCCAAGCTGAAGACATTTTAAACTATGCGCCATTCAT 664
 QY 515 AlaPheThrIleLysThrValLysGluAsnLeuAspGlnLeuLysAsnGlnIleGln 534
 Db 665 GCATTTACTATTAATAAGCTGTAAAGAAACCTTGATCATCTGGAACCAATAACAGG 724
 QY 535 ProCysLysLysPheAlaIleAlaAspAlaThrArgGluAspProPheLysGluLys 554
 Db 725 CCATCGAGAAAGTGGCGCT-GCGGATGCGAC-CGAGAGATCCCTTTAAAGGAA 779
 QY 555 LeuGluIleMetThrArgIleGlnThrTyrCysGlnMetSerPro-MetSerAspIleG 574
 Db 780 ---TGTGAATATGGCAGGCTTCCACTTGGTGCAGT---AGTCCAGGTGCGA---TTGGG 830
 QY 574 YThrGlnProTyrGluGlnTyrAlaIleGlnMetGluLysLysAlaAlaLysLysGly 594
 Db 831 TAGCCACCTATTAATA-----GGGGCTCAACGGGGGAAAAAGGGGCTAAAGAGTAA 884
 QY 594 nArgLysGluArgValCysAlaGlnHisLeuArgLysTyrAsnGluAlaLeuGlnHis 614
 Db 885 ACGGTTGGGCGACCTTGGGCGCAACAGCGCACAAAGAGACAGAGAGAGAGTAA 944
 QY 614 nAspThrIleArgMetIleAspAlaTyrThrHis 625
 Db 945 CGAACACAGCAAA-----GACAGAAACGACACAC 972

RESULT	5
LOCUS	Bf98j236
DEFINITION	Bf98j236 729 bp mRNA linear EST 24-JAN-2001 60230587JF1 NIH_MGC_88 Homo sapiens cDNA IMAGE:4397083.5.
ACCESSION	Bf98j236
VERSION	Bf98j236.1
KEYWORDS	GI:12386048
SOURCE	EST. human.
ORGANISM	Homo sapiens
REFERENCE	Eumayr et al., Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 729)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
TITLE	Contact: Robert Strausberg, Ph.D. Email: csapbs@email.nih.gov
JOURNAL	Tissue Procurement: ATCC
COMMENT	cDNA Library Preparation: Life Technologies, Inc. CDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://limage.llnl.gov Plate: Llam10097 row: a column: 20 High quality sequence stop: 665. Location/Qualifiers 1..729 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4397083." /clone_lib="NIH_MGC_88" /tissue_type="Duodenal adenocarcinoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: small intestine; Vector: pMW-SPOUT6; site:1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.767 kb. library enriched for full-length clones and constructed by life technologies. Note: This is a NIH-MGC Library."
BASE COUNT	240 A 137 C 184 G 168 T
ORIGIN	
Alignment Scores:	
Pred. No.: 1,06e-115 Length: 729	
Score: 1129.00 Matches: 238	
Percent Similarity: 98.36% Conservative: 2	
Best Local Similarity: 97.54% Mismatch: 4	
Query Match: 21.26% Indels: 3	
DB: 12 Gaps: 0	
US-09-515-363C-2 (1-1025) x Bf98j236 (1-729)	
OY	140 GIUGUIGULeuLeuThrllegLUsApRygsAnArqllleaAlaaAagluAsnaSnqly 159
Dd	1 GAGGAGAACACTGTTCACATTTGAACAAGAAACC GGATTGCTGCAGAAAACCAATGCA 60
OY	160 AsngUsnerGlVValARgLUeuLeuuysArglllevA.JLnLyglUsnLtppPheSer 179
Dd	61 AA TGATA TCAGGGTGTAAGACAGCTACTCAA AAAAGCAT TTGTGCCAGAAAAAC CTGTTCT 120
OY	180 AlAPhelueAsnValLeuaRGlnthrGrlyasnasngLuLeuValGInguLeuThrly 199
Dd	121 GCATTTC GAATCGTCTTC GTCCA AACAGSAA ACAMTAGAA TTGTGTCCAAACGTTAAC AGCAC 180
OY	200 SerAsPyCsSrergLUserASnalaelutlleGuasnLeuSERGINvalAspyCyProGln 219
Dd	181 ICTGTATTGCTCGAAGAAGCAATGCAGAGATTGAGAAATT TAACAAAGTTATGTCCTCA 240
OY	220 ValGUguIgUlneuleuSe rThrhValGLInPrOASNLeugLUlsygLIuValTrpCl y 239
Dd	241 CTCGAAAGACCAACTCTTCACCAACAGCTTCAGCCC AATCTGTGAGAAAAGAGGTCTGGGRC 300

OY	240	MetsGuanaaaanSerSerGcUuseSerPheAlaaspSerSeValValSerUouaaasp	259			
DB	301	ATGGGAATAAAGTCAATCGAAATCATCTTTTGACGATTCTTGAGACTTCATAAATCAAC	360			
OY	250	ThurSerLeuaIagIUGISerValSerCySLeuaSpGIuSerLeugIyHIsenaaAsn	279			
DB	361	ACAATTTGGCGAAGAAGGTCTCAGCTGCTTAGATGAAAGCTTGAGCATACACAAAC	420			
OY	280	MetGlYSerASPserGlyThrMetGlYSerASPserASPglUGluasnaValaaLArg	299			
DB	421	ATGGGCAGAGATTCCAGGCCACCATGGGAAGTCAATCGAATGAAGAAGAAATGGTAAGA	480			
OY	300	AlaSerProGUuProGUleuGUleuArpGofyrGImetGluValAlaSeruuAla	319			
DB	481	GCAATCCCCGGAGCCAGAACTCCAGCTCAGGCTTACCAAATGGAAAGTTGCCASDAAAC-	539			
OY	320	LeugIUGlUGlYSasnIellellecySLeuPProTrfIGlySerGlyLSrThra-gyaAla	339			
DB	540	TTCGAAAGGAAACAATATCATCATCTGCGTC-CCTACAGGAGAGTGGAAAAACCAAGAGT	598			
OY	340	VAlTYrIIealalyaSPHisLeuaSPlysLySLysAlaSerGIuProTrfILysVal	359			
DB	599	GTTTCATGTGCCAAGATCATCTTACACAG-AAGAAAAAAGCATCTGAG-CTCGAAAATT	657			
OY	360	IleValIleValaSnLysValLeuLeuValIgUGInLeuPheArgLysGIuHeh-Tro	379			
DB	658	ATAGTCTTGTCATAAAGSTACTGCTAGTTGTAACAGCTCTTCGAAAAGAGTTCACATA	717			
OY	380	PheLeuLysLys	383			
DB	718	TTTTTGGAGAAAG	729			
RESULT 6						
AKO18602						
LOCUS						
DEFINITION	AKO18602	1197 bp	mRNA linear hit# 19-JAN-2002			
ACCESION	Mus musculus adult male cecum CDNA, RIKEN full-length enriched library, clone:9130009C22:homoIog to METANOOMA DIFFERENTIATION ASSOCIATED PROTEIN-5, full insert sequence.					
VERSION	AKO18602.1 GI:12858393					
KEYWORDS	HTC; CAP trapper.					
SOURCE	Mus musculus (strain:C57BL/6J) adult male cecum cDNA to mRNA, clone:libRIKEN full-length enriched mouse cDNA library clone:9130009C22.					
ORGANISM						
Mus musculus						
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.						
REFERENCE						
1 Carninci,P. and Hayashizaki,Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999)						
JOURNAL MEDLINE	99279253					
PUBMED	10349636					
REFERENCE						
2 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)						
JOURNAL MEDLINE	20499374					
PUBMED	11042159					
REFERENCE						
3 Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Kono,H., Akiyama,J., Nishi,K., Kitsumai,T., Tashiro,H., Itoh,M., Suna,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,I., Kasliwalgi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsubara,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kiru,A. and Hayashizaki,Y. Riken integrated sequence analysis (RISA) system--484 format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000)						
JOURNAL						

QY 982 LeuLysLeuArgAspPheValValPheLysAspAspSerThrLysLysGlnIleLys 1001
 DB 603 CTTAAATAGAGATTTTGTAGTCAATTTCAAAATACCTACCCAGAGAACAGTAGACAG 662
 QY 1002 TystPrpValGluLeuProIleThrPheProAsnLeuAspTyrSerGluTyrScyLeuPhe 1021
 DB 663 AAGTGGTGGAAATGGCTTTCAGATTTCTGTATCTGACTACTCAAAATACCTGCTGTAT 722
 QY 1022 SerAspGluAsp 1025
 DB 723 AGTGATGAAGAT 734
 RESULT 7
 BG741146
 LOCUS BG741146 827 bp mRNA linear EST 15-MAY-2001
 DEFINITION 602631817p1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4776869 5',
 mRNA sequence.
 ACCESSION BG741146
 VERSION BG741146.1 GI:14051799
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 827)
 NIH-MGC http://mgs.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 AUTHORS Unpublished (1999)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cga@bbs-remail.nih.gov
 Tissue Procurement: James Cleaver, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA
 Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L14M10630 row: b column: 06
 High quality sequence stop: 763.
 FEATURES
 source location/Qualifiers
 1..827
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4776869"
 /clone_lib="NCI_CGAP_Skn3"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: skin; Vector: pCMV-SPORE; Site_1: NCI;
 Site_2: Salt; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.5kb. Library constructed by Life
 Technologies. Note: this is a NCI-CGAP library."
 BASE COUNT 254 a 186 c 204 g 183 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 3,93e-112 Length: 827
 Score: 1098.00 Matches: 255
 Percent Similarity: 91.17% Conservative: 3
 Best Local Similarity: 90.11% Mismatches: 13
 Query Match: 20.67% Indels: 15
 DB: 12 Gaps: 0
 US-09-515-363c-2 (1-1025) x BG741146 (1-827)
 QY 87 GlySerProLeuAlaAlaArgTyrMetAsnProGluLeuThrAspLeuProSerProSer 106
 DB 2 GGCAGCCCTCTGGCCGC-CGCTACATGAACCCGAGCTCAGAGCTGCTCTCCATCG 60
 QY 107 PhcGluAsnAlaHisProGluTyrLeuGlnLeuLeuAsnLeuGln-f:ohrLeuVal 126
 DB 61 TTTGAGAACGCTCATATGATATTCACCACTCTGACCTCCTCAGTCTCAGTCTGCT 120
 QY 126 LasPlysLeuValArgAspValLeuAspLysCysMetGluGluGlnLeuThrTrl 146

DB 121 GGACAGCTTCTAGTTAGAGAGCTCTTGATTAAGTGCATGGAGAGAACTGTTCAT 180
 QY 146 eGluAspArgAsnArgIleAlaAlaGluAsnAsnGlySerGlyValArg 166
 DB 181 TGAACAGCAAAACCGATTCGTGTCGAGAAACCAATGAAATGAAATGAGTGAAGA 240
 QY 166 uLeuLeuysArgIleValGlnLysGluAsnTrpPheSerAlaPheLeuAsnValGln 186
 DB 241 GCTACTAAAGAGATTGCGCAAGAAAGAAACGTTCTCTGATTTCTGAAATGTTTCG 300
 QY 186 gGlnThrGlyAsnAsnGluLeuValGlnGluLeuThrGlySerAsp-CysSerGlnSer 205
 DB 301 TCAACAGCAAAACCAATGAACTGTTCAGAGATTACAGGCTCTGATGGCTCAAAAAGC 360
 QY 206 AsnAlaGluIleGluAsnLeuSerGlnValAspGlyProGlnValGluGlnLeu 225
 DB 361 AATGACAGATGAGAAATTTATCACAGTTGATGCTCTCAAGTGAAGAGCACTTCT 420
 QY 226 SerThrThrValGlnProAsnLeuGlu-LysGluValTrpGlyMetGluAsnAsnSerSe 245
 DB 421 TCACACAGCTTCAGCCAAATCTGGACAGAGAGGCTGGGCGATGGAAATACCTATC 480
 QY 245 rGluSerSerPheAla-AspSerSerValSerGluSerAspThrSerLeuVal 264
 DB 481 ACAATCATCTTTTGCACCATTTCTGTAGTTTCAGAAATCAGACAAAGTTTCTCA 540
 QY 264 uGlySerValSerCysLeuAspGluSerLeuGly-HisAsnSerAsnMetGlnSerAsp 284
 DB 541 AGCAAGTGCACCTCTTATGATGAAGCTTGGACCATACGCAACATGGCTAGTAT 600
 QY 284 eGluThrMetGlySerAspSerAspGlu-GluAsnValAlaAlaArgAlaSerProGlu 303
 DB 601 CAGGACCATGGGAAGTATTCAGATGAACGAAATGGGACGACAAAGCATCCCTGAG 660
 QY 304 ProGluLeuGlnLeuArgProTyrGlnMet-GluValAlaGlnProAlaLeuGluGly 323
 DB 661 CCAGAACTCCACCTCAGGCGCTTACCAATGGCAGAGTTGCCAGGCAGC-TGGCCAA 719
 QY 323 yAsnIleIleIleCysLeuProThrGlySerGlyLysThrArgValAlaValTyrIle 342
 DB 720 ACACATATCATCATCTGCTC-CCTACAGGAGGAGGAAAAACAAAGCTGTGATTCAT 778
 QY 343 AlaLysAspHisLeuAspLysLysLysLysAlaSerGluProGly 357
 DB 779 GCCAGGATCCTTAGACAGAGAAACAAAGCATCTGAGCCTGGA 823
 RESULT 8
 BF686405 781 bp mRNA linear EST 22-DEC-2000
 LOCUS BF686405 602143786p1 NIH-MGC_46 Homo sapiens cDNA clone IMAGE:474805 5',
 mRNA sequence.
 DEFINITION 602143786p1 NIH-MGC_46 Homo sapiens cDNA clone IMAGE:474805 5',
 mRNA sequence.
 ACCESSION BF686405
 VERSION BF686405.1 GI:11971813
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 781)
 NIH-MGC http://mgs.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 AUTHORS Unpublished (1999)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cga@bbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L14M1170 row: 1 column: 22

QY 625 sLeuGIuThrPheTYrAsnGIuLysAspLys-LysPheAlaValIleG nAspAspse 645
 |||||
 Db 122 TCTTGAACCTTCTAATAGAGAGAGAAAGATAGAAAGTTTGCAGTCATCAAGATGATAG 181
 QY 645 rAspGIuGIyLysAspAspGIuTYrCysAspGIyAspGIyAspGIyAspGIyLys 665
 |||||
 Db 182 TCATGAGCGTGTGTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 241
 QY 665 sProLeuLysLeuAspGIuThrAspAspPheLeuMetThrLeuPhePheC nAspAsp 685
 |||||
 Db 242 ACCCTTGAACCTGATGAGAGAGAGATAGATTTCTCATGCTTATTTTGGAAACCAATTA 301
 QY 685 sMetLeuLysAspLeuAlaGIuAsnProGIuTYrGIuAsnGIuLysLeuThrLysLeuAr 705
 |||||
 Db 302 AATGTGTAAGAAAGCGTGGCTGAAAGCCCAACATATGAAATGAAAGCTGATCAATTAAG 361
 QY 705 gAsnThrIleMetGIuGIuTYrThrArgThrGIuGIuSerAlaArgGIyT eIlePheTh 725
 |||||
 Db 362 AATATCATTAATGAGCAATATAGTACTAGAGTACGAGATGACGAGAGATTAATCTTAC 421
 QY 725 rLysThrArgGIuSerAlaTYrAlaLeuSerGIuThrPheThrGIuAsnGIuLysPheAl 745
 |||||
 Db 422 AAAAACACACAGAGTGCATATGCGCTTCCAGTGGATTACTGAAATGAAATTTTC 481
 QY 745 aGIuValGIyValLysAlaAlaHisLeuIleGIyAlaGIyHisSerSerC nPheLysPr 765
 |||||
 Db 482 TGAAGTAGAGAGAGTCAAAAGCCCAACATCTGATTTGAGTGCACAGAGAGTCTTCAAC 541
 QY 765 oMetThrGIuAsnGIuGIuLysGIuValIleSerLysPheArgThrGIyLys-LysAsp 785
 |||||
 Db 542 CATGACACAGAAATGAAACAAAGAGAGTCAATTAATTTCCAGTCAATTAATTAAT 601
 QY 785 eLeuLeuIleAlaThrThrAlaAlaGIuGIuLysGIuLysGIuLysGIuLysGIuLys 805
 |||||
 Db 602 TCGTTATTCCTTACCCAGAGTGCAGAGAGAGTCCGATATTCACGACTGTACC--ACGC 658
 QY 805 lArGIyGIyLysValIleThrAsnGIuIleAlaMetValGIuAlaArgGIyArgAlaArg 825
 |||||
 Db 659 TATCCGACAGCGCCAGCCACCATGATAGCCATGAGC-CAGCCCGCGGCTCAAGCCAAAC 717
 QY 825 lAspGIuSerThrTYrValLysValAlaHisSerGIySerGIyValIleGIuHisGI 845
 |||||
 Db 718 TACTGAAAGAGACCAAAAGACGCGCTGCCCAAAAGGCTCA--CAGCCACA 765
 QY 845 hTyAlaAsnAspPheArgGIuLysMetMetTYrLysAlaIleHisCysValGIuAsnMet 865
 |||||
 Db 766 CAACCCAC--GCGACACACCGCCACACCATCCCA 798
 QY 865 ySProGIuGIuTYrAlaHisLysIleLeuGIuLysIleMetGIuSerIlePheGIuLys 885
 |||||
 Db 799 AAACACAGCATGCACACCAACCCCAACATGCG-- 830
 QY 885 ySMeLysThrLysArgAsnIleAlaLysHisTYrLysAsnAsnProSer 901
 |||||
 Db 831 -----GCCACACACTATCAACCCAGCCGAGC 857
 RESULT 10
 BM476961 1046 bp mRNA lines EST 05-FEB-2002
 LOCUS BM476961
 DEFINITION AGEMCOURT_6481569 NIH_MGC_71 Homo sapiens cDNA: one IMAGE:555523
 5' mRNA sequence.
 ACCESSION BM476961
 VERSION BM476961.1 GI:18526003
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 1046)
 TITLE NIH-MGC http://mhc.nci.nih.gov/
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs.rem@nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM2275 row: n column: 04
 High quality sequence stop: 622.
 Location/Qualifiers
 source
 1..1046
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:555523"
 /clone_lib="NIH-MGC_71"
 /rclone_type="telomoyosarcoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site: 1; Notif:
 Site 2: Salt: Cloned unidirectionally. Primer: 5'-GTGAGT
 http://image.llnl.gov
 Average insert size 2.1 kb.
 BASE COUNT 318 a 217 c 227 g 282 t 2 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,27e-108 Length: 1046
 Score: 1066.00 Matches: 209
 Percent Similarity: 98.14% Conservative: 2
 Best Local Similarity: 97.21% Mismatches: 4
 Query Match: 20.07% Indels: 0
 DB: 13 Gaps: 0
 US-09-515-363C-2 (1-1025) x BM476961 (1-1046)
 QY 2 SerAsnGIyTYrSerThrAspGIuAsnPheArgTYrLeuIleSerCysPheAlaArg 21
 |||||
 Db 1 TGAATGGGTATTCACAGACAGAGAAATTCGCTATCATCTGCTTCAAGG 60
 QY 22 ValLysMetTYrIleGIuValGIuProValLeuAspTYrLeuThrPheLeuProAla 41
 |||||
 Db 61 GTGAATAATGATATCAGGTGAGGCTGTGCTGAGTCACTGACCTTCTGTCTGTGTGAG 120
 QY 42 ValLysGIuGIuIleGIuArgThrValAlaThrSerGIyAsnMetGIuAlaGIuLeu 61
 |||||
 Db 121 CTGAAGAGACAGATTCAGAGACAGACTCCGACCTCCGGACATCAGCAGCTTGAAG 180
 QY 62 LeuLeuSerThrLeuGIuLysGIyValIlePheLysGIyTYrPheArgGIuPheVal 81
 |||||
 Db 181 CTGCTGAGACCTTGCAGAGAGAGAGTGGCAGCTTGGTGGACTCGGGAATTCGTGAG 240
 QY 82 AlaLeuArgArgThrGIySerProLeuAlaAlaArgTYrMetAsnProGIuLysPheAla 101
 |||||
 Db 241 GCCCTCCGAGACCGGAGGAGCCCTGCGCCGCTCATGAAACCTGAGTCAACGAG 300
 QY 102 LeuProSerProSerPheGIuAsnAlaHisAspGIyTYrLeuGIuLeuLysAsnLeu 121
 |||||
 Db 301 TTGCCCTCTCCATCGCTTTGAGAACCTCATGATGAATCTCCAACTGCTGACCTCT 360
 QY 122 GIuProThrLeuValAspLysLeuLeuValArgAspValLeuAspLysCysMetGIu 141
 |||||
 Db 361 CAGCCCACTCTGTGTGACAGCTTCTAGTAGAGAGCGTCTTGGAATGATGATGAGAG 420
 QY 142 GIuLeuLeuThrIleGIuAspArgAsnArgIleAlaAlaGIuAsnAsnGIyAsnGI 161
 |||||
 Db 421 GAACCTGTGCAATATGAGAGACAGAAACCGCATTTCTGCTGCAGAAACAAATGAGAA 480
 QY 162 SerGIyValArgGIuLeuLeuLysArgIleValGIuLysGIuAsnProPheSerAla 181
 |||||
 Db 481 TCAGGTGTAGAGAGCTCTATAAAGAGATGTGGAGAAAGAAATGGTCTCTGCAATT 540
 QY 182 LeuAsnValLeuArgIleThrGIyAsnAsnGIuLeuValGIuGIuLeuThrCysArg 201
 |||||
 Db 541 CTGAATGTCTTCTGTCAACAGAGAAACAAATGAACCTTGCCAAAGATTAAAGGCTTGA 600


```

/db_xref="taxon:9606"
/clone="IMAGE:5532660"
/clone_lib="NH_MGC_71"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: uterus; Vector: pCMV-Sport5; Site: 1; Note:
Site: 2; Salt: Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.1 kb."
BASE COUNT      393 a      285 c      243 g      318 t
ORIGIN
Alignment Scores:
Pred. No.:      4,53e-107      Length:      1239
Score:          1055.50      Matches:      229
Percent Similarity: 86.91%      Conservative: 10
Best Local Similarity: 83.27%      Mismatches: 23
Query Match:    19.87%      Indels:      13
DB:              Gaps:      4
US-09-515-363c-2 (1-1025) x BM467774 (1-1239)
QY      190 AsnAsnGluLeuValGlnGluLeuThrGlySerAspCysSerGlnSerAlaGluIle 209
      |||||
DB      4 AACAAATGAAGCTGCTCCAGAGTTAAACAGGCTGATGCTCAGAAAGCAATGCAGAGATT 63
QY      210 GlnAsnLeuSerGlnValAspGlyProGlnValGlnGluLeuLeuSerThrThrVal 229
      |||||
DB      64 GAAATATTATCAACAATGATGCTCTCAAGTGAAGAGCAACTTCTTTTACCAACAGATT 123
QY      230 GlnProAsnLeuGluLysGluValTyrGlyMetGlnAsnAsnSerSerGlnSerSerPhe 249
      |||||
DB      124 CAGCCAAATCTGAGAGAGAGGCTCTGGGCATGAGAAATACACATCAGATCATCTTTT 183
QY      250 AlaAspSerSerValValSerGlnSerAspThrSerLeuAlaGluLysValSerGln 269
      |||||
DB      184 GCAGATCTCTCTGATCTTACAGAAATGAGACACAAATTTGACAGAGAAATGCAGCTTC 243
QY      270 LeuAsnGluSerLeuGlnHisAsnSerAsnMetGlySerAspSerGlyThrMetGlySer 289
      |||||
DB      244 TTAGATGAAGAGCTTGTGACATTAACACACATGCGAGTATTCAGGCAATCATG36AAGT 303
QY      290 AspSerAspGluGlnAsnValAlaAlaArgAlaSerProGlnProGlnGlnLeuVal 309
      |||||
DB      304 GATTCAATGTAAGAGAAATGGGAGCAGACAGCATCCCGAGGACAAATCTCAGCTCATG 363
QY      310 ProTyrGlnMetGluValAlaGlnProAlaLeuGluGlyLysAsnIleIleIleCysLeu 329
      |||||
DB      364 CCTTACCAATGAGAGATTGCGCCAGCCGCTTGAAGAGAGAAATATCATCATCTGCTTC 423
QY      330 ProThrGlySerGlyLysThrArgValAlaValTyrIleAlaLysAspHisLeuAspLys 349
      |||||
DB      424 CCAACAGGAGGTGAAAAACAGAGAGGCTGTTACATTGCCAAGATCTCATGACAG 483
QY      350 LysLysLysAlaSerGlnProGlyLysValIleValLeuValAsnLysValLeuLeuVal 369
      |||||
DB      484 AAGAAAAAGCATCTGAGCGCTGAAAAAGTTATAGCTCTTTCATTAAGTACAGCAAGTT 543
QY      370 GlnGlnLeuPheArgLysGluPheGlnPro-PheLeuLysLysTyrPheArgValIleGln 389
      |||||
DB      544 GAACAGCTCTCTCCGAGAGAGTCCCAACCATTTTGAAGAAAGGACATCTGTTATTCG 603
QY      389 LysLeuSer-GlyAspThrGlnLeuLysIleSerPheProGlnValValLys-SerCysAsp 408
      |||||
DB      604 ATTAAAGGGGTATACCAACATGAAATATCATTTCCAGAGTTGCAATTCCTGTCAT 663
QY      409 IleIle-----IleSerThrAlaGlnIleLeuGlnLysSer-LeuLeuAsnLeu---GI 425
      |||||
DB      664 ATTATATATCACTACAGCTCAAAATTCCTTTAAAAAAGCTCCGCCCTTAAGCTTGGAAA 723
QY      425 uaaanglygluasp-----AlaGlyValGlnLeuSerAspPheSerLeuIleIleLys 443
      |||||
DB      724 AATGGGAAAAAAGCGCTTGGCGGCTTCCAACTGCCAAGAAATTTTCCCTTCAT----- 778

```

```

QY      443 pGluCysHisHisThrAsnLysGluAlaValTyrAsn 455
      |||||
DB      779 -----TATCATTTGATGAAGAAATGTTATCATCCACAC 808
      |||||
RESULT 13
LOCUS      BQ316075/c
DEFINITION BQ316075
ACCESSION BQ316075
VERSION    BQ316075.1
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Hominidae; Homo.
REFERENCE  1 (bases 1 to 672)
            Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Bittencourt,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
            Brunstein,A., deoliveira,P.S., Bucher,P., Jomeneel,V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S. and
            Simpson,A.J.G.
            Shotgun sequencing of the human transcriptome with cDNA expressed
            sequence tags
            Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
            Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-3704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?tbl=CM3&tbl2=CM3-CT0275-
            191099-024-g99&tbl3=1999-10-19&tbl4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 50
            High quality sequence stop: 671.
            Location/Qualifiers
                1..672
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_lib="CT0275"
                /dev_stage="Adult"
                /note="Organ: colon; Vector: puc18; Site: 1; Sma1; Site: 2;
                Sma1; A mini-library was made by cloning products derived
                from ORESTES PCR (U.S. Letters Patent application No. 196
                716 - Ludwig Institute for Cancer Research) profiles
                into the puc 18 vector. Reverse transcription of tissue
                mRNA and cDNA amplification were performed under low
                stringency conditions."
BASE COUNT      165 a      158 c      120 g      229 t
ORIGIN
Alignment Scores:
Pred. No.:      4,46e-103      Length:      672
Score:          1016.00      Matches:      202
Percent Similarity: 94.62%      Conservative: 9
Best Local Similarity: 90.58%      Mismatches: 10
Query Match:    19.13%      Indels:      2
DB:              Gaps:      2
US-09-515-363c-2 (1-1025) x BQ316075 (1-672)
QY      652 GlnTyrCysAspGlyAspGluAspAspLeuLysProLeuLysLysPheVal 671
      |||||
DB      671 GAGTATGCGATGCGATGATGAGATGATTTAAAGAAACCTTTGAACTGAGAGAA 612
      |||||
QY      672 ThrAspArgPheLeuMetThrLeuPheGlnGluAsnAsnLysMetLeuLysAla 691
      |||||

```



```

OY 61 LLLSTLEKGVHMLGWTREVEALRRTGSPPLAARYNNPELTDLPSPSPENHDEYEDLNL 120
DB 61 LLLSTLEKGVHMLGWTREVEALRRTGSPPLAARYNNPELTDLPSPSPENHDEYEDLNL 120
OY 121 LQPTLVYDKLLVROVLDKCKMEBELLTIEDNRNIAAENNGNESVPELTK-VOKEREFSA 180
DB 121 LQPTLVYDKLLVROVLDKCKMEBELLTIEDNRNIAAENNGNESVPELTK-VOKEREFSA 180
OY 181 FLNVLROGTGNMELVOELTSDGSCSENAELENLSQVDPQVEEQLSTTVYLNLEKIVWGM 240
DB 181 FLNVLROGTGNMELVOELTSDGSCSENAELENLSQVDPQVEEQLSTTVYLNLEKIVWGM 240
OY 241 ENNSSESSFADSSVYSESDTSLAEGSVSCDSESLGHSNMGSSGNGMS-SUDEENVAARA 300
DB 241 ENNSSESSFADSSVYSESDTSLAEGSVSCDSESLGHSNMGSSGNGMS-SUDEENVAARA 300
OY 301 SPEPELOLRPYOMEVAQPALEGKNIIICLPDTSCKTRAVAYIAKDHLDKKKASEPKVY 360
DB 301 SPEPELOLRPYOMEVAQPALEGKNIIICLPDTSCKTRAVAYIAKDHLDKKKASEPKVY 360
OY 361 VLVNKKVLLVEOLFRKEFOPFLKKMYRYIGLSGDTQKISPEVYKSCDII-STAQILENS 420
DB 361 VLVNKKVLLVEOLFRKEFOPFLKKMYRYIGLSGDTQKISPEVYKSCDII-STAQILENS 420
OY 421 LNLNENGEDAGVOLSDFSLIIDCHHTNKEAVYNNIMRYILMKKKKN-1KKENFPVJP 480
DB 421 LNLNENGEDAGVOLSDFSLIIDCHHTNKEAVYNNIMRYILMKKKKN-1KKENFPVJP 480
OY 481 LPQILGLTASPGVGATQAKAEHILKLCANLDAFTIKIVKNLIDOLK-VJOEPCPKFA 540
DB 481 LPQILGLTASPGVGATQAKAEHILKLCANLDAFTIKIVKNLIDOLK-VJOEPCPKFA 540
OY 541 IADATREDPFKEKLEIMTRIQTYCOMSPMSDFGTPYEWALOMEKKAACKGRKEKVC 600
DB 541 IADATREDPFKEKLEIMTRIQTYCOMSPMSDFGTPYEWALOMEKKAACKGRKEKVC 600
OY 601 AEHLRKYNEALQINDTIRIMDAYTHLETFYNEEKDKKFAVIEDDSDEG-DEYCGDDHDE 660
DB 601 AEHLRKYNEALQINDTIRIMDAYTHLETFYNEEKDKKFAVIEDDSDEG-DEYCGDDHDE 660
OY 661 DDLKPKLIDETDRFLMTLEFFENNMKMLKLAENPEYENKLTKLANTIM-VYTRTEESAR 720
DB 661 DDLKPKLIDETDRFLMTLEFFENNMKMLKLAENPEYENKLTKLANTIM-VYTRTEESAR 720
OY 721 GIIFTRTOSAVYALSOITENEFKFAVGVYKAHNLIGAGHSEFKEMTON-VQEVISKPEY 780
DB 721 GIIFTRTOSAVYALSOITENEFKFAVGVYKAHNLIGAGHSEFKEMTON-VQEVISKPEY 780
OY 781 GKINLILATTVAEGLDIEKCNVIRYGLVTNFIAMVQAGARADESTVYLVANSGCV 840
DB 781 GKINLILATTVAEGLDIEKCNVIRYGLVTNFIAMVQAGARADESTVYLVANSGCV 840
OY 841 IEHETVNDERKKMYKALICVOYNNMKPEYAHNLELOMOSIMEKKKTKKNLAKHKNP 900
DB 841 IEHETVNDERKKMYKALICVOYNNMKPEYAHNLELOMOSIMEKKKTKKNLAKHKNP 900
OY 901 SLITTELCCKMSVYACGEBDIHYIEKHHVNNTPREKELYVRENKALOKK-ADYOINGET 960
DB 901 SLITTELCCKMSVYACGEBDIHYIEKHHVNNTPREKELYVRENKALOKK-ADYOINGET 960
OY 961 ICGGAGMGTMMVHKGDLDPCLKIRNFVYVFNKNSTKKQKKMVELPITFIMLDYSECC 1020
DB 961 ICGGAGMGTMMVHKGDLDPCLKIRNFVYVFNKNSTKKQKKMVELPITFIMLDYSECC 1020
OY 1021 FSDSD 1025
DB 1021 FSDSD 1025

```

```

AC 09BYX4:
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE RNA helicase-DEAD box protein RH116.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Coccone C., Kolesnichenko V., Billaut-Mulot O., Tuomola M.-J.,
RA Capron A., Bahr G.M.;
RT "Identification of a new RNA helicase (RH116) regulated by the
RT Immunomodulator Murabutide."
RL Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AY017378; AAG54076.1;
DR InterPro: IPR001315; CARD.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF00270; DEAD_1.
DR Pfam: PF00271; Helicase_C_1.
DR SMART: SM00487; DEXDC_1.
DR SMART: SM00490; HELICG_1.
DR PROSITE: PS50209; CARD_1.
KW ATP-binding; Helicase.
SQ SEQUENCE 1025 AA; 116671 MW; 6B4F3D14E88A7D6D CRC64;

Query Match          99.5%; Score 5285; DB 4; Length 1025;
Best Local Similarity 99.5%; Pred. No. 4,9e-287;
Matches 1020; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 MSGNYSTDENFYVLI SCFRAVYKMYIOVEPVLDVLTFLPAEYKEQIQRIATSCNMKAVE 60
DB 1 MSGNYSTDENFYVLI SCFRAVYKMYIOVEPVLDVLTFLPAEYKEQIQRIATSCNMKAVE 60
OY 61 LLLSTLEKGVHMLGWTREVEALRRTGSPPLAARYNNPELTDLPSPSPENHDEYEDLNL 120
DB 61 LLLSTLEKGVHMLGWTREVEALRRTGSPPLAARYNNPELTDLPSPSPENHDEYEDLNL 120
OY 121 LQPTLVYDKLLVROVLDKCKMEBELLTIEDNRNIAAENNGNESVPELTK-VOKEREFSA 180
DB 121 LQPTLVYDKLLVROVLDKCKMEBELLTIEDNRNIAAENNGNESVPELTK-VOKEREFSA 180
OY 181 FLNVLROGTGNMELVOELTSDGSCSENAELENLSQVDPQVEEQLSTTVYLNLEKIVWGM 240
DB 181 FLNVLROGTGNMELVOELTSDGSCSENAELENLSQVDPQVEEQLSTTVYLNLEKIVWGM 240
OY 241 ENNSSESSFADSSVYSESDTSLAEGSVSCDSESLGHSNMGSSGNGMS-SUDEENVAARA 300
DB 241 ENNSSESSFADSSVYSESDTSLAEGSVSCDSESLGHSNMGSSGNGMS-SUDEENVAARA 300
OY 301 SPEPELOLRPYOMEVAQPALEGKNIIICLPDTSCKTRAVAYIAKDHLDKKKASEPKVY 360
DB 301 SPEPELOLRPYOMEVAQPALEGKNIIICLPDTSCKTRAVAYIAKDHLDKKKASEPKVY 360
OY 361 VLVNKKVLLVEOLFRKEFOPFLKKMYRYIGLSGDTQKISPEVYKSCDII-STAQILENS 420
DB 361 VLVNKKVLLVEOLFRKEFOPFLKKMYRYIGLSGDTQKISPEVYKSCDII-STAQILENS 420
OY 421 LNLNENGEDAGVOLSDFSLIIDCHHTNKEAVYNNIMRYILMKKKKN-1KKENFPVJP 480
DB 421 LNLNENGEDAGVOLSDFSLIIDCHHTNKEAVYNNIMRYILMKKKKN-1KKENFPVJP 480
OY 481 LPQILGLTASPGVGATQAKAEHILKLCANLDAFTIKIVKNLIDOLK-VJOEPCPKFA 540
DB 481 LPQILGLTASPGVGATQAKAEHILKLCANLDAFTIKIVKNLIDOLK-VJOEPCPKFA 540
OY 541 IADATREDPFKEKLEIMTRIQTYCOMSPMSDFGTPYEWALOMEKKAACKGRKEKVC 600
DB 541 IADATREDPFKEKLEIMTRIQTYCOMSPMSDFGTPYEWALOMEKKAACKGRKEKVC 600
OY 601 AEHLRKYNEALQINDTIRIMDAYTHLETFYNEEKDKKFAVIEDDSDEGDDHY-DEGDHDE 660

```

```

Db 601 AEHLRKYNAKALQINDTIRMIDAYTHLETFFYNEKKKFAVIEDDSDEGSHLEYCGDEJE 660
QY 661 DDLAKKPKLDETRFLMTLFFENNNKMLKRLAENPEYENKLTLYLNTIMEQYTRIESAR 720
Db 661 DDLAKKPKLDETRFLMTLFFENNNKMLKRLAENPEYENKLTLYLNTIMEQYTRIESAR 720
QY 721 GILFTKROSAYALSOMITENEFKAEYGVKAHHLIGAGHSEKPMPTONQKQVISEKFI 780
Db 721 GILFTKROSAYALSOMITENEFKAEYGVKAHHLIGAGHSEKPMPTONQKQVISEKFI 780
QY 781 GKINLLIATVAEGLDICEKNIVIRYGLVTNEIAMVOAGRAADESTYVLVAHSOSGV 840
Db 781 GKINLLIATVAEGLDICEKNIVIRYGLVTNEIAMVOAGRAADESTYVLVAHSOSGV 840
QY 841 IEHETVNDPEKKMYKAIRHVOVNMKPEYAKHLLLEQMSIMEKKMKTAKNIAKHKNP 900
Db 841 IEHETVNDPEKKMYKAIRHVOVNMKPEYAKHLLLEQMSIMEKKMKTAKNIAKHKNP 900
QY 901 SLITFLCKNCVSLACSGEDIHVEIKMHVNMTEPEKELYVRENKALOKKCAQYQINGEL 960
Db 901 SLITFLCKNCVSLACSGEDIHVEIKMHVNMTEPEKELYVRENKALOKKCAQYQINGEL 960
QY 961 ICKGQAMGTMMVHKGLDPLCKLRNFVVFKNNSKTKQYKAWELPITFPLNDSYSCCL 1020
Db 961 ICKGQAMGTMMVHKGLDPLCKLRNFVVFKNNSKTKQYKAWELPITFPLNDSYSCCL 1020
QY 1021 FSDSD 1025
Db 1021 FSDSD 1025

RESULT 3
08R5F7 PRELIMINARY; PRT; 1025 AA.
ID 08R5F7;
AC 08R5F7;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE HELICARD.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN (1)
RA SEQUENCE FROM N.A.
RT Kovacsovich M., Hofmann K., Tschopp J.;
RT "HELICARD, a novel CARD-containing helicase."
RT Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY075132; AAC77205.1; -.
SQ SEQUENCE 1025 AA; 115971 MW; 708FCAC690CFF6B8 CMC61;

Query Match 79.1%; Score 4199; DB 11; Length 1025;
Best Local Similarity 79.7%; Pred. No. 2.2e-22;
Matches 818; Conservative 75; Mismatches 131; Indels 2; Gaps 2;

```

```

Db 241 DDLPLASCTDSSVTSIESDTSLAEGSVSCFDESLGNSNMKGDSCGMSDSESVLYQIKR 300
QY 300 ASPEPELQRPYOMEYAQALBGNIIICLPFGSGKTRAAVIAADHLDKKASERAKV 359
Db 301 VSPEPELQRPYOMEYAQALBGNIIICLPFGSGKTRAAVIAADHLDKKASERAKV 360
QY 360 IYLVNVLVEOLFREPEPPLKMYRVIGLSGDPOLKISPEPVKSCDIIISTOJLEN 419
Db 361 IYLVNVLVEOLFREPEPPLKMYRVIGLSGDPOLKISPEPVKSCDIIISTOJLEN 420
QY 420 SLINLENGEDAGVQSDPSLIIDECNHTNKSAVYNNIMRHLYMOKLNKNNRLKKNRPVY 479
Db 421 SLINLENGEDAGVQSDPSLIIDECNHTNKSAVYNNIMRHLYMOKLNKNNRLKKNRPVY 480
QY 480 PLPQIIGLTASPGVCAATQAKAEHILKLCANLDAFTIKYKENLDLKNQIOEPCKE 539
Db 481 PLPQIIGLTASPGVCAATQAKAEHILKLCANLDAFTIKYKENLDLKNQIOEPCKE 540
QY 540 ATADATREDPFEKLEIEMTRIOYQMSPMDSFGTOPYEQWAIOMEKKAARKRELY 599
Db 541 VIADOTREMPFEKLEIMASIOYQMSPMDSFGTOPYEQWAIOMEKKAARKRELY 600
QY 600 CAEHLRYNEALQINDTIRMIDAYTHLETFFYNEKKKFAVIEDDSDEGDEYCDGDEP 659
Db 601 CAEHLRYNEALQINDTIRMIDAYTHLETFFYNEKKKFAVIEDDSDEGDEYCDGDEP 659
QY 660 EDDLAKKPKLDETRFLMTLFFENNNKMLKRLAENPEYENKLTLYLNTIMEQYTRIESA 719
Db 660 EDDLAKKPKLDETRFLMTLFFENNNKMLKRLAENPEYENKLTLYLNTIMEQYTRIESA 719
QY 719 RGIIFTKROSAYALSOMITENEFKAEYGVKAHHLIGAGHSEKPMPTONQKQVISEKFI 779
Db 720 RGIIFTKROSAYALSOMITENEFKAEYGVKAHHLIGAGHSEKPMPTONQKQVISEKFI 779
QY 779 TGRINLLIATVAEGLDICEKNIVIRYGLVTNEIAMVOAGRAADESTYVLVAHSOSGV 839
Db 780 TGRINLLIATVAEGLDICEKNIVIRYGLVTNEIAMVOAGRAADESTYVLVAHSOSGV 839
QY 839 VIEHETVNDPEKKMYKAIRHVOVNMKPEYAKHLLLEQMSIMEKKMKTAKNIAKHKNP 899
Db 840 VIEHETVNDPEKKMYKAIRHVOVNMKPEYAKHLLLEQMSIMEKKMKTAKNIAKHKNP 899
QY 899 PSILITFLCKNCVSLACSGEDIHVEIKMHVNMTEPEKELYVRENKALOKKCAQYQINGEL 959
Db 900 PSILITFLCKNCVSLACSGEDIHVEIKMHVNMTEPEKELYVRENKALOKKCAQYQINGEL 959
QY 959 IICKGQAMGTMMVHKGLDPLCKLRNFVVFKNNSKTKQYKAWELPITFPLNDSYSCCL 1019
Db 960 IICKGQAMGTMMVHKGLDPLCKLRNFVVFKNNSKTKQYKAWELPITFPLNDSYSCCL 1019
QY 1020 FSDSD 1025
Db 1020 FSDSD 1025

RESULT 4
096MX8 PRELIMINARY; PRT; 468 AA.
ID 096MX8;
AC 096MX8;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE CDNA FLJ31731 f15, clone NTZRT2006855, weakly similar to Homo sapiens
DE RNA helicase (HUG-I) mRNA.
OS Homo sapiens (Homo).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RA SEQUENCE FROM N.A.
RT Ishibashi T., Kanehori K., Yosida M., Matanabe S., Ishida S., No Y.,
RT Houta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J.,

```

RA Matanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
 RA Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,
 RA Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
 RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
 RA Magatsuna M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
 RA Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isojima T.,
 RT *NPOD human cDNA sequencing project. *
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK056293; BAB71141.1; -.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR001650; Helicase_C.
 DR Pfam: PF00271; Helicase_C; 1.
 DR ATP-binding; Helicase.
 SQ SEQUENCE 468 AA: 54470 MW: 330405507436DFA CRC64:
 Query Match 46.2%; Score 2456; DB 4; Length 468;
 Best Local Similarity 99.1%; Pred. No. 4e-127;
 Matches 464; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 558 MTRIQTCOMSPMSDFGTPYEQMAIOMEKKAAGKGRKRCVCAEHLKRYNEALQINDTI 617
 DB 1 MTRIQTCOMSPMSDFGTPYEQMAIOMEKKAAGKGRKRCVCAEHLKRYNEALQINDTI 60
 QY 618 RMIDATHTLETFYNEKKKKFAVIEDSDGDDGDEYCDDEDEDLKRLKIDETIRELM 677
 DB 61 RMIDATHTLETFYNEKKKKFAVIEDSDGDDGDEYCDDEDEDLKRLKIDETIRELM 120
 QY 678 TLEFENNMKRLKLAENPEYENKTLRLNTIMEOYTRTESAGIIFTKIKOSAYALSOW 737
 DB 121 TLEFENNMKRLKLAENPEYENKTLRLNTIMEOYTRTESAGIIFTKIKOSAYALSOW 180
 QY 738 ITNEKFAEYGVAAHHLIAGHSSEKPTONEOKEVISKFRITKINLITVAEGLD 797
 DB 181 ITNEKFAEYGVAAHHLIAGHSSEKPTONEOKEVISKFRITKINLITVAEGLD 240
 QY 798 IKECNIVIRGLVTNEIAVQARGARADESTYLVLAHSGSGVIEHETVNDFREKMYKA 857
 DB 241 IKECNIVIRGLVTNEIAVQARGARADESTYLVLAHSGSGVIEHETVNDFREKMYKA 300
 QY 858 IHCVMNKKDEYAAHKLLELOMOSIMEKKMKTKRNIAKHKNNSLITFLKNCVYLAGSG 917
 DB 301 IHCVMNKKDEYAAHKLLELOMOSIMEKKMKTKRNIAKHKNNSLITFLKNCVYLAGSG 360
 QY 918 EDIHVEKKNHVNTPREFELYIVRENKALOKKCAOYUINGELICQGAQMGTMVNHKL 977
 DB 361 EDIHVEKKNHVNTPREFELYIVRENKALOKKCAOYUINGELICQGAQMGTMVNHKL 420
 QY 978 DLPCILIRNFVYVFNKNSTKQYKKWVELPTFPNLDVSECCIFSDSD 1025
 DB 421 DLPCILIRNFVYVFNKNSTKQYKKWVELPTFPNLDVSECCIFSDSD 1468
 RESULT 5
 Q99KS4 PRELIMINARY: PRT: 514 AA.
 AC 099KS4:
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Hypothetical 59.4 kDa protein (Fragment).
 GN 913009G22RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC04031; AAH04031.1; -.
 DR MGD; MGI:191836; 913009G22RIK.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR001650; Helicase_C.

DR Pfam: PF00271; Helicase_C; 1.
 DR SMART; SM00490; HELICE; 1.
 KW ATP-binding; Helicase; Hypothetical protein.
 FT NON-TER 1
 SQ SEQUENCE 514 AA: 59429 MW: 9025A02BE680F1F CRC64:
 Query Match 42.8%; Score 2271.5; DB 11; Length 514;
 Best Local Similarity 85.0%; Pred. No. 6.1e-117;
 Matches 438; Conservative 30; Mismatches 46; Indels 1; Gaps 1;
 QY 511 ANLDAFTIKTYKENIDOLKNOIOEPCKKFAIADATREPPFEKLEIMTRIQTCOMSPM 570
 DB 1 ANLDAFTIKTYKENIDOLKNOIOEPCKKFAIADATREPPFEKLEIMTRIQTCOMSPM 60
 QY 571 SDGTPYEQMAIOMEKKAAGKGRKRCVCAEHLKRYNEALQINDTIIDAHTLETFY 630
 DB 61 SDGTPYEQMAIOMEKKAAGKGRKRCVCAEHLKRYNEALQINDTIIDAHTLETFY 120
 QY 631 NEKKKKFAVIEDSDGDDGDEYCDDEDEDLKRLKIDETIRELMITFENNMKRL 690
 DB 121 TDEKERRFAVL-NDSDSDDEASSCNDLKGQVKSILKIDETIRELMITFENNMKRL 179
 QY 691 AENPEYENKTLRLNTIMEOYTRTESAGIIFTKIKOSAYALSOWITNEKFAEYGV 750
 DB 180 AENPEYENKTLRLNTIMEOYTRTESAGIIFTKIKOSAYALSOWITNEKFAEYGV 239
 QY 751 AHHLIAGHSSEKPTONEOKEVISKFRITKINLITVAEGLDICEKNIVIRGLV 810
 DB 240 AHHLIAGHSSEKPTONEOKEVISKFRITKINLITVAEGLDICEKNIVIRGLV 299
 QY 811 TNEIAVQARGARADESTYLVLAHSGSGVIEHETVNDFREKMYKAICHVONKKPREYA 870
 DB 300 TNEIAVQARGARADESTYLVLAHSGSGVIEHETVNDFREKMYKAICHVONKKPREYA 359
 QY 871 KHLLELOMOSIMEKKMKTKRNIAKHKNNSLITFLKNCVYLAGSGEDIVIEKMHVN 930
 DB 360 KHLLELOMOSIMEKKMKTKRNIAKHKNNSLITFLKNCVYLAGSGEDIVIEKMHVN 419
 QY 931 MTRPEKELYVRENKALOKKCAOYUINGELICQGAQMGTMVNHKGLDPLCLIRNFVYV 990
 DB 420 MTRPEKELYVRENKALOKKCAOYUINGELICQGAQMGTMVNHKGLDPLCLIRNFVYV 479
 QY 991 FKNSTKQYKKWVELPTFPNLDVSECCIFSDSD 1025
 DB 480 FKNSTKQYKKWVELPTFPNLDVSECCIFSDSD 514
 RESULT 6
 Q8R144 PRELIMINARY: PRT: 467 AA.
 AC 08R144:
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical 54.0 kDa protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC025508; AAH25508.1; -.
 DR Hypothetical protein.
 SQ SEQUENCE 467 AA: 53966 MW: F3B0D976778F0442 CRC64:
 Query Match 38.8%; Score 2059.5; DB 11; Length 467;
 Best Local Similarity 84.6%; Pred. No. 2.4e-105;
 Matches 396; Conservative 28; Mismatches 43; Indels 1; Gaps 1;
 QY 558 MTRIQTCOMSPMSDFGTPYEQMAIOMEKKAAGKGRKRCVCAEHLKRYNEALQINDTI 617
 DB 1 MTRIQTCOMSPMSDFGTPYEQMAIOMEKKAAGKGRKRCVCAEHLKRYNEALQINDTI 60

```

Db 1 MASIQYCOKSPKSDFGTGYEOMAIOMERKAKADGNRRCVCAEHLRKNALQINDI 60
OY 618 RMIDATHTLETPENEKDKKPAVEDDSDEGGDEYCDGJEDDOLKRLKLEIDTRCLM 677
Db 61 RMIDASHLETFTDEKKEKFAVL-NDSESDDEASSCMLQKGVKKSILKDEIDELIM 119
OY 678 TLEFENKMLKRLAENPEYENKELTKLRNTIMEOYTRIEASAMILEKIKOSAYALSNOM 737
Db 120 NLEFQKKMLKRLAENPEYENKELTKLRNTIMEOYTRIEASAMILEKIKOSAYALSNOM 179
OY 738 ITENEKFAEYGVKAHHLIGAGHSSEKPYMTONEQEVISKFRIGKTNLLIATVAESID 797
Db 180 IMENAFAYGVKAHHLIGAGHSSEKPYMTONEQEVISKFRIGKTNLLIATVAESID 239
OY 798 IKCNIVIRGLVTNEIAYVQAGRARADESTVLAHSVIEHEETVNDPREKKMYKA 857
Db 240 IKCNIVIRGLVTNEIAYVQAGRARADESTVLAHSVIEHEETVNDPREKKMYKA 299
OY 858 INCVQMKPEEYNAHKLLELOMOSIMEKKTKENIAKHKNPSLITELKNCVLAESG 917
Db 300 INCVQMKPEEYNAHKLLELOMOSIMEKKTKENIAKHKNPSLITELKNCVLAESG 359
OY 918 EDIHVEKMHVNMTPPEFKELTYVRENKALQKKCADYQNGELTCKGQAMGTIMVHKCL 977
Db 360 ENIHVEKMHVNMTPPEFKELTYVRENKALQKKCADYQNGELTCKGQAMGTIMVHKCL 419
OY 978 DLPECLIRNFVYVKNSTKKQYKKNVELPITPBNLDYS-SCCLFSESD 1025
Db 420 DLPECLIRNFVYVKNSTKKQYKKNVELPITPBNLDYS-SCCLFSESD 167

RESULT 7
OY 099J87 PRELIMINARY: PRT: 678 AA.
AC 099J87;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Hypothetical 76.7 kDa protein.
GN DILGPZE OR DILGPZ.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed-1161808;
RA Miyoshi K., Cui Y., Riedinger G., Lehoczeky J., von L., Oka T.,
RT "Structure of the Mouse Stat 3/5 Locus: Evolution from Protophila to
RT Zebrafish to Mouse."
RL Genomics 71:150-155(2001).
DR EMBL; AF317000; AAK15475.1; -
DR EMBL; AF316999; AAK15474.1; -
DR MGD; MG1:1931560; DILGPZE.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00270; DEAD. 1.
DR Pfam; PF00271; helicase_C. 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
KW ATP-binding; Helicase; Hypothetical protein.
SQ
SEQUENCE 678 AA: 76726 MW: DCA2B75A3AD376A8 JRC64:

Query Match 25.5%; Score 1356; DB 11; Length 678;
Best Local Similarity 41.7%; Pred. No. 1.7e+66;
Matches 299; Conservative 129; Mismatches 237; Indels 52; Gaps 12;
OY 306 IOIRPYOMEYAPALGKNNIICLPGSGKTRAVYIAKHLLKKKKASIPGVYLVNK 365
Db 1 MELRPOMEYIPLALGKNNIICLPGSGKTRAAVAVAKRHLD---TVKGVVAVLVNK 56
OY 366 VLLVEOLFKEFOPFLKKWYRIGLSGTOLKISPEVAVKSCILISTALLENLNLIE 425

```

```

Db 57 VHLVSO-HAEFRRLDKHWVTTLTSGMGSAAGGLMARSHDLICTAELIQT V NESSE 115
OY 426 NQEDAGVQLSDPSLLIIDECHTNEEAYNNIMRTYLMOKLKNLKENK VHLI POL 485
Db 116 --EDERVELREPSLIVDECHHTKDYNTLLISRYLEOKK----KAE-----PQVL 164
OY 486 GLTASPGVGAATKQAKAEHILKLCANLDAFTIKVRENLDLQKNGQPCAKFA ADAT 545
Db 165 GLTASPGVGAATKQAKADHILQLCANLDTCHIMSPKCYSLMLMNRKPCQVYV-CORR 224
OY 546 REDPEKLELIMTRIGYQMSPM-SDFGTQPYBOWAIOMERKAKGNRAEYV ABEL 604
Db 225 AADPEGDLIKKLMNQIHQOLEMPDLKQOGTQMEYQVYVQLCKDAEAGLDEQFYALHL 284
OY 605 KRYNALQINDITRMIDATYHLETYNEKOKKPAVEDDSDEGGDEYCDGJEDDOLK 664
Db 285 RRYNDALFIHDYVRAADALMLQDEYDRERTKTOMVAES----- 325
OY 665 KRLKLEIDTRLEMTLEFENKMLKRLAENPEYENKELTKLRNTIMEOYTRIEASAMILE 724
Db 326 -----WLLKLEFDKKNVGLQLAARG-ENKLEKLERILKQF-GSHGHIGIF 373
OY 725 TKTROSAYALSNOMITENEKFAEYGVKAHHLIGAGHSSEKPYMTONEQEVISKFRIGKIN 784
Db 374 TKTROSAYALSNOMITENEKFAEYGVKAHHLIGAGHSSEKPYMTONEQEVISKFRIGKIN 433
OY 785 LLIATVAESGIDIECNIVIRGLVTNEIAYVQAGRARADESTVLAHSVIEHE 844
Db 434 LLIATVAESGIDIECNIVIRGLVTNEIAYVQAGRARADESTVLAHSVIEHE 493
OY 845 TVNDPREKKMYKALHCYQMKPEEYNAHKLLELOMOSIMEKKTKENIAKHKNPSLIT 904
Db 494 LTNEALEVMEKAVAAVQMDPEFRKAKIRDOQASLVKRAARAAREIQQGFLREHVQ 553
OY 905 FLCKNCSVLASGEDIHYIEKMHVNMTPPEFKELTYVRENKALQKKCADYQNGELTCK 962
Db 554 LDCINOMVAVGSGDLRKKEGHHVNMTPPEFKELTYVRENKALQKKCADYQNGELTCK 613
OY 963 KCGQAMGTIMVHKCLDLPCLIRNFVYVKNSTKKQYKKNVELPITPBNLDYS-SCCL 1019
Db 614 NCGEWMGFQMIYKSVTLPLVKIGS--MLEETPRGKIQAKMSRVFSPFVETLQMG 668

RESULT 8
OY 096C10 PRELIMINARY: PRT: 678 AA.
AC 096C10;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Hypothetical 76.6 kDa protein.
GN Homo sapiens (Human).
OS Homo sapiens; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Eukaryota; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE-COLON;
RA Strausberg R.;
RT Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC014949; AAH14949.1; -
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00270; DEAD. 1.
DR Pfam; PF00271; helicase_C. 1.
KW ATP-binding; Helicase; Hypothetical protein.
SQ
SEQUENCE 678 AA: 76612 MW: 859E1749C731D06 CRC64:

Query Match 25.5%; Score 1352; DB 4; Length 678;
Best Local Similarity 40.7%; Pred. No. 2.8e+66;
Matches 292; Conservative 136; Mismatches 237; Indels 52; Gaps 12;

```

OY 306 LQLRPYMEVAOPALEGNNIIICLPFGSGKTRVAVYIAADHLDKKAASKPCKVIVLVNK 365
 DB 1 MELRSYOMEVIMPALEGGNNIIIMLPFGAGKTRAAVYAKRHLE---TV---AKVYLVNR 56
 OY 366 VLLVEQLFRKREPPLKKMYRVIGSGDGLKISPEVVKSCDIIISTWILKNSLNLNLE 425
 DB 57 VHLVYQ--HGEERRRMLDGRWYVTTLSGDMGPRAGFGHARCHDLICTAELLQMALTSPE 115
 OY 426 NGEADVQLSDPSLIIIDECHTNKEAVYNNIMRHYLMOKLNNKLNKKNRPVIRPLDVL 485
 DB 116 --EEHVELTVSLIVDECHTNKDYVNNVMSQYLELALQRAQY-----PLQVYL 164
 OY 486 GLTASPGVGGATKQAKAEENHILKICANDAFITKTYKENIDQLKNOIOEPCKKRAIADAT 545
 DB 165 GLTASPGVGGATKQAKAEENHILKICANDAFITKTYKENIDQLKNOIOEPCKKRAIADAT 224
 OY 546 REDPPEKLEIMTRIGTQCOMSPNS--DEGTOPYEOMALOMEKKAAGNKKERVCAPHL 604
 DB 225 SODPFGDLKTKMDQIHNLHLEPELSRKFGTOMEDQVYKALSEAALAGLOBORYALHL 284
 OY 605 RRYNEALQINDTIRMIDATYHLETFYNEEKDKKFAVIEDSDSDDEVDDEDEDLX 664
 DB 285 RRYNDALLIHDIVRAVDALALADQFYHREHVTKTQIL----- 322
 OY 665 KPLKIDETDFRLMTFFENNNKMLKLAENPEYENKLTGKRLNTMEQYTFEESAKRIIF 724
 DB 323 -----ARRRLALFDKRNKELALATNGP--ENKLEMLKLTGROFS--SSNSPRLIF 373
 OY 725 TKTROSAVALSOWITENENKFAEVGKANHILIGAGSSEFKPTONOEKVEISKRTKIN 784
 DB 374 TKTROSAHSLIMLLOQOGLQTVDIRAOLLIGAGSSOSTHMTQROQEVIOKPOQILN 433
 OY 785 LLATTTVAEBGDIKECNIVIRGLVTNEIAMVQARGRADESTYVYLVASGSGVIFHE 844
 DB 434 LLVATTSVAEBGDIKPCNVVVRGGLTNEISVQARGRAKADOSYAVAFALDGSRELKRE 493
 OY 845 TVNDFREKMYKAIHGVONMKPREYAKHILELOMOSIMEKKMKTRNIAKHNKNNPILIT 904
 DB 494 LINEALETLMEQVAAVQAKMDQAEYQAKTRDLOQALALTKRAAOAORENKKOOPVEHVO 553
 OY 905 FLCKNCVSLACSGEDIHVIEKMHVNMTPPEKELY--IYREKKALOKKCAVOINGEILCK 963
 DB 554 LLCTINCMVAVGHSGLRKVEGTHHVNPNFSYTVNSKRPVYINKVFKKRPQGVLSNR 613
 OY 964 -CGOANGTMVYHKGDLPLKLRNFVYVFKNNSTKKQYKQWVELPTFPNDYSECC 1019
 DB 614 NCGEVMGLQMYKSKVLPLVKVRS--MLETFQGRIOAKKMSRVFSPVDFDLOHC 668
 RESULT 9
 O9HAM6 PRELIMINARY: PRT: 678 AA.
 AC 09HAM6
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DB Hypothetical 76.6 kDa protein.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=EMBRYO;
 RA Iisogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
 Tanase T., Komura Y., Togiya S., Komai F., Hara K., Takuchi K.,
 Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
 Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y., Oshima A.,
 NEO human cDNA sequencing project.
 RT Submitted (AUG-2000) to the EMBL/GenBank/DDBI databases.
 RL EMBL: AK021416; BAB13818.1;
 DR InterPro: IPR001410; DEAD.

DR InterPro: IPR001650; Helicase_C.
 DR Pfam: PF00270; DEAD. 1.
 DR Pfam: PF00271; Helicase_C. 1.
 DR SMART: SM00487; DEXC. 1.
 DR SMART: SM00490; HELIC. 1.
 KM ATP-binding; Helicase.
 SO SEQUENCE 678 AA; 76642 MW; 0628A21A71856E5 CRC64;
 Query Match 25.3%; Score 1344; DB 4; Length 678;
 Best Local Similarity 40.6%; Pred. No. 7.7e-66;
 Matches 291; Conservative 136; Mismatches 238; Indels 52; Gaps 12;
 OY 306 LQLRPYMEVAOPALEGNNIIICLPFGSGKTRVAVYIAADHLDKKAASKPCKVIVLVNK 365
 DB 1 MELRSYOMEVIMPALEGGNNIIIMLPFGAGKTRAAVYAKRHLE---TV---AKVYLVNR 56
 OY 366 VLLVEQLFRKREPPLKKMYRVIGSGDGLKISPEVVKSCDIIISTWILKNSLNLNLE 425
 DB 57 VHLVYQ--HGEERRRMLDGRWYVTTLSGDMGPRAGFGHARCHDLICTAELLQMALTSPE 115
 OY 426 NGEADVQLSDPSLIIIDECHTNKEAVYNNIMRHYLMOKLNNKLNKKNRPVIRPLDVL 485
 DB 116 --EEHVELTVSLIVDECHTNKDYVNNVMSQYLELALQRAQY-----PLQVYL 164
 OY 486 GLTASPGVGGATKQAKAEENHILKICANDAFITKTYKENIDQLKNOIOEPCKKRAIADAT 545
 DB 165 GLTASPGVGGATKQAKAEENHILKICANDAFITKTYKENIDQLKNOIOEPCKKRAIADAT 224
 OY 546 REDPPEKLEIMTRIGTQCOMSPNS--DEGTOPYEOMALOMEKKAAGNKKERVCAPHL 604
 DB 225 SODPFGDLKTKMDQIHNLHLEPELSRKFGTOMEDQVYKALSEAALAGLOBORYALHL 284
 OY 605 RRYNEALQINDTIRMIDATYHLETFYNEEKDKKFAVIEDSDSDDEVDDEDEDLX 664
 DB 285 RRYNDALLIHDIVRAVDALALADQFYHREHVTKTQIL-----C----- 322
 OY 665 KPLKIDETDFRLMTFFENNNKMLKLAENPEYENKLTGKRLNTMEQYTFEESAKRIIF 724
 DB 323 -----ARRRLALFDKRNKELALATNGP--ENKLEMLKLTGROFS--SSNSPRLIF 373
 OY 725 TKTROSAVALSOWITENENKFAEVGKANHILIGAGSSEFKPTONOEKVEISKRTKIN 784
 DB 374 TKTROSAHSLIMLLOQOGLQTVDIRAOLLIGAGSSOSTHMTQROQEVIOKPOQILN 433
 OY 785 LLATTTVAEBGDIKECNIVIRGLVTNEIAMVQARGRADESTYVYLVASGSGVIFHE 844
 DB 434 LLVATTSVAEBGDIKPCNVVVRGGLTNEISVQARGRAKADOSYAVAFALDGSRELKRE 493
 OY 845 TVNDFREKMYKAIHGVONMKPREYAKHILELOMOSIMEKKMKTRNIAKHNKNNPILIT 904
 DB 494 LINEALETLMEQVAAVQAKMDQAEYQAKTRDLOQALALTKRAAOAORENKKOOPVEHVO 553
 OY 905 FLCKNCVSLACSGEDIHVIEKMHVNMTPPEKELY--IYREKKALOKKCAVOINGEILCK 963
 DB 554 LLCTINCMVAVGHSGLRKVEGTHHVNPNFSYTVNSKRPVYINKVFKKRPQGVLSNR 613
 OY 964 -CGOANGTMVYHKGDLPLKLRNFVYVFKNNSTKKQYKQWVELPTFPNDYSECC 1019
 DB 614 NCGEVMGLQMYKSKVLPLVKVRS--MLETFQGRIOAKKMSRVFSPVDFDLOHC 668
 RESULT 10
 O95786 PRELIMINARY: PRT: 925 AA.
 AC 095786
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DB RNA helicase.
 GN RIG-I.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Sun Y.W.;
 RT "Rig-I", a human homolog gene of RNA helicase, is induced by retinoic acid during the differentiation of acute promyelocytic leukemia cell."
 RL Thesis (1997), Shanghai Institute of Hematology, Rui-Jin Hospital, Shanghai Second Medical University.
 RN (2)
 RP SEQUENCE FROM N.A.
 RA Yi-Wu S.;
 RL Submitted (DEC-1997) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF039963; AAD19826.1;
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR001650; Helicase_C.
 DR Pfam: PF00270; DEAD; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR SMART: SM00490; HELIC; 1.
 DR ATP-binding: Helicase.
 KM SEQUENCE 925 AA; 106613 MW; 5922BAF3DD0F00D4 CRC64;
 SO
 Query Match 20.5%; Score 1086.5; DB 4; Length 925;
 Best Local Similarity 31.2%; Pred. No. 1,6e-51;
 Matches 325; Conservative 166; Mismatches 164; Indels 185; Gaps 35;

QY 12 RLILSCFRAVVKYIOVEPYLDLT--FLPAVKEQIQIVATSGMVAQVELLSTLEKG 69
 DB 6 RSLQAFODYIRKTLPTLYLSTYMAFWFREYV-QYIAIKNNKGPEAAFLFLKLE 63
 QY 70 VMHIGTREFEVALRRTGSPFLAAYMNPETDLPSPFENAH--DEYLOLINTLOPTLYVD 127
 DB 64 LOEGMGRGLADLDIAG--YSGLYEAEISMD-----FKIKLELEYRLIKLOLPEFK 116
 QY 128 KLLVDYLOKMEELLETIEDNRNIAAENNGESVRELKRIYQ--K-NMFSAPLNL 185
 DB 117 RIIPDTIIDL-LSQCLINCEBELOLQSTKGMMAEKIIECLISDKIMPKTL---- 171
 QY 186 RGTGNNELVQELGSDGSENAELENLSQVDPQVEQLSTIVQPLKELVGMEN--- 242
 DB 172 -----KLALEKERNKFSELMIVKGIKVDETEDLDMKTSIQ-----IFYQDEPCT 219
 QY 243 -NSSESFADSVYSESDTSLAGSVSCDSESLGHNSNMQSDSGTMSDNIENVAARAS 301
 DB 220 QNLSENGCPSEV---SDTL-----YS 239
 QY 302 PEPELOLRPYQMEVAQPALEGKNIILCLPTSGKTRAVAYIAKDHLKKKASEPGKVIY 361
 DB 240 P--FKFRNYQLDELALPAMKGNITICAPYCGCKTFVSLICHHL-KKILQGGKGVVF 295
 QY 362 LVKKVLLVEQ--LFRKEPPLKKMRYVIGLSDTOLKLSFEVAKSLIISFAILE 418
 DB 296 FANQIDPYEQKSVFSKYFE--RHGFRVIGISGATAENPVQIYENNI--ILIPQILV 352
 QY 419 NSLNLNENGDAVQSLDFSLIITDECHRTNKAENVYNIWRHYLMQKLNKRLKKEKVIY 476
 DB 353 N---NLKKGITP--SLSTIFLMTFDECHNTSKQHPYIMLNPNTLDOKLQKLS--SS----- 399
 QY 479 IPLPQILGLTASPGVCGATQAAAEHILKLANLAFITKTYKENIDQIKNOIDEPCK 538
 DB 400 GPLPQVIGLTASVGVCAKMTDALDYICKLASLDSAVIATVKHNELEI FQVYKQKQK 459
 QY 539 FATADATREDPFE-----KLEIETRIQTYCQSPKAPNRTQTPKQW 581
 DB 460 FKRVESHISDKFYIILAQIMRDESLAKRICKLEMLNSLQIN-----FLPQIKLQW 532
 QY 582 AIDMEKKA-----AKGKRKERV-----AEHLKRYEALQINDITPMIAYHLEIYN 631
 DB 513 IIVYQAKCMFQMPDKDESRICKALPLTYSHLRKYDALIISEHAMKI ALDYLAQTES 572
 QY 632 EKKDKRAVIEDSDGDEYDQGDDEDEDLKPLKLTEDRFLTLFFENKMKLKLAA 691

DB 573 NVNAGFEEIED-----LTPREEMJELHSVS 601
 QY 692 ENPEYENKLTLMRTIMEQYRTTESARGIIFETKTRQSAVALSQWENEA--AFVGV 749
 DB 602 RQPSNENPKLEDLCFLQDEYHINPETIT-LLVYTRALVYDLKMMIESNLSIAQD 660
 QY 750 KAHNHLGAGHSEFPKPTQNEQKEVYSKPR-TGKINLLATTAEGDLKKNVIVHYG 808
 DB 661 ----LGRCKTQNTQNTMTLPKQKCLIDAFKAGDINILATISVADGIDIAQVNLVHYE 716
 QY 809 LVTNEIAYQAGRARADESTYLVLAHSGSVIENETVNDPEKKMKYKAIRVYMKPEE 868
 DB 717 YGVNVIKMIQTGRGRARSKCFLLT-SNAGVIEKQIMYKKEKMMNSILKCIYDEAV 775
 QY 869 YAKHLELOMOSIMEKKMKTRKNIKHY--KNNPSLITFLCKNSVLAQSGHIVIEK 926
 DB 776 FRKLIHQTH--EKFINDSQEKPPVPDKENKKL--LCKKCALACTYMAVAVIEE 829
 QY 927 HHVNMTPREKELYIENKALQKCAQYQINCEIIC--KCGQAMTMMVHKGILDPCK 983
 DB 830 HTVLDQAFKEQFVSRRPK-PKQSSFEKRAKIFCARQNGSHDMGJHVKYKIFELPYIK 888
 QY 984 IRNEVYVFKNNSTKKQYKRW 1003
 DB 889 IESFVEDIATGVOTLYSKW 908
 RESULT 11
 ID 09GLV6 PRELIMINARY; PRT; 940 AA.
 AC 09GLV6;
 DT 01-MAR-2001 (TREMBLrel, 16, Created)
 DT 01-MAR-2001 (TREMBLrel, 16, last sequence update)
 DT 01-DEC-2001 (TREMBLrel, 19, last annotation update)
 DE RNA helicase.
 GN RHIV-1.
 OS Sus scrofa (Pig).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA MEDLINE-20261798; PubMed-10799277;
 RX Zhang X., Wang C., Schook L.B., Hawken R.J., Rutherford M.S.;
 RT "An RNA helicase, RHIV-1, induced by porcine reproduction and
 RT respiratory syndrome virus (PRRSV) is mapped on porcine chromosome
 RT 10q13."
 RL Microb. Pathog. 28:267-278(2000).
 DR EMBL: AF181119; AAG09428.1;
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR001650; Helicase_C.
 DR Pfam: PF00270; DEAD; 1.
 DR Pfam: PF00271; helicase_C; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR SMART: SM00490; HELIC; 1.
 DR ATP-binding: Helicase.
 SO SEQUENCE 940 AA; 107583 MW; 118CA910B0AF7821 CRC64;
 Query Match 19.2%; Score 1018; DB 6; Length 940;
 Best Local Similarity 30.1%; Pred. No. 9,6e-48;
 Matches 315; Conservative 158; Mismatches 369; Indels 204; Gaps 35;

QY 31 VLDYLT--FLPAVKEQIQIVATSGMVAQVELLSTLEKGMHIGTREFEVALRRTGS 88
 DB 25 ILSYMAPWFRDDEV-OHIOEKNNKGPTEASLFLQFLLE-LOEGMGRGLADLDIAG 81
 QY 89 PLARYMNPETLD-LPSPFENAH--DEYLOLINTLOPTLYVDKLVYDVLKQWME----- 140
 DB 82 -----YCG--LCAISLWMPFKTEKLEEVRSILRLQPEFKTINPKDLFI IAH LLSQ 134
 QY 141 --BELLTIEDNRNIAAENNGESVRELKRIYQKENFSAFLNVLROTGNNEI VOELT 198
 DB 135 ECEELIOLICSSKQLMA-----GAKEVCEILRS--DKENN----- 167

OY 199 GSDGSESNALINLSQVDPQVEBOLLSTTVQNLNLEKVMGNNSSTSPADSSVSES 258
 DB 168 -----PKTLKLALEKE-----ESRSELMWVKGAEVYK 196
 OY 259 DTSIAGSGVCLDESIGH-----NSNNGSDGTGMSDSDENVAARASPHPELQRP--Y 311
 DB 197 MKLEDEDMKTCVQIYFKKEPENONL-----NONGSSAPHTYPLKPRKY 244
 OY 312 QMEVAOPALGSKNITITLPRTGSGKTRAVYIAMDHLDKKKASEPGKVIVLVKVLVEQ 371
 DB 245 QLELALPAQCKNTIICAPGCKTFFVSLICEHHL-KKIPRGRKGVYHIALQIVYRQ 303
 OY 372 ---LFRKEFPPLKMWRYVGLSGDQGLKISPEVYKSCDIIISTANGLNLSLNLENCE 428
 DB 304 QKVSFKNHE---RLGYKAGISGASIDVQVQIVENSQITITPOLANCLTN----- 395
 OY 429 DAGV--OLSDPSLIIIDECHTNKKEAVYNNIMRHYLMQKKNRKLKNNAVPLPCQIIO 486
 DB 356 ---GTPSLSVFTLMIPECHNTSKQHPYVIMSYLDRIAGSS-----DSLQVQIG 405
 OY 487 LTAPOVGAATKQAKAEHLKLCANIDAFITVKNENLQKNOIOLPKKPAIDATK 546
 DB 406 LTAPOVGAATKQAKAEHLKLCASIDTSVATVRDNLLELREYVYKPKPRKVELKT 465
 OY 547 EDPFKKLEIMTRIQYQ-----MSPHSDTQTPYEDYAIOMEKKA-- 590
 DB 466 TDRKCTISOLMEITSLASISEELQITTLGLFOIUNSNFQTOYEDMIVKQKEVY 525
 OY 591 ---KKNRKRKRC-----AEHLKRYNEALQINDTRIMPAVYHLEFVNEKDKKFAVY 641
 DB 526 FQMPDKDKESRICKALFSYMSHLRIYNDALIMHAKMKALDYLDQIFPNIRAAFG--- 582
 OY 642 EDDSDGDEYCOGDEDEDLKKPLKLDTRFLMTLFFENKMKRLAENPEYENKI 701
 DB 583 -----DEIEQDL-----TQR-----FREKQLELESISIDPSNENPKI 614
 OY 702 TKLRNTIMEQYTRTEESARGIIFTKTROSAYALSOMITENKRAVYVANHILGCHSS 761
 DB 615 RDLCTILOEYHNLNPE--KILLEVKTALVDALKMKIKNPUS--FLKPSITLGCKTN 671
 OY 762 EFPNTOEKOEVISKFRIGIN-LIATVAEGLDIKKCNIVIRYGVITNEIAVQAR 820
 DB 672 QNGMTPAQKCVLDFRTDKDKILITTSVADEGIDIVAGCNVILEYVQVNTKMIQIR 791
 OY 821 GRARADESTVYLAHSGSGVIEHETVNDPREKMYKAIHCVQNMKDEFAVHKLLEOMOS 880
 DB 732 GGRGRAGSKCFLLT--ANADLIDKEKMMYKEENMGAILILQTMDEAVFPRKIHOQI-- 788
 OY 881 IMKKKKTKRNIAKHYKKNNSUL-----TFLLCKNSVLAQSGEDHIVFPMHIVMTEPE 934
 DB 789 -----REKILTRONOGKREPVPDKTKKLLCKKCAKATVADIRVNEKCHFTVVGNA 840
 OY 935 FKELATVIRENKALOKKCAQOINGELIIC--KCGQAMGTMMVHKSJDLATKIRNVVVF 991
 DB 841 FREREFYSKLHPK-PKSPGNIEKAKIYCARPDCSHDMGLIVYKAFEMPRKLESTVVD 899
 OY 992 KNNSTKKQYKKWELPITPNDLYSE 1017
 DB 900 IATGVQTVAHAKMDENFEKLSFDAE 925
 RESULT 12
 OY9D25 PRELIMINARY: PRT: 210 AA.
 AC Q9D25;
 DT 01-JUN-2001 (TREMblrel. 17, Created)
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
 DE 9130009C22RLK protein.
 GN 9130009C22RLK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN (1)
 RC SEQUENCE FROM N.A.
 RX STRAIN=C57BL/6J; TISSUE=CECUM;
 MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shigaawa A., Shibata K., Yoshino M., Itoh M., Ishi Y.,
 Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Ikuoka S.,
 Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
 Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 Fleischmann M., Gaasterland T., Gissi C., King B., Kochia H.,
 Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 Schiml L.M., Staudil F., Suzuki R., Tomita M., Wagner L., Watanabe T.,
 Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barish B.,
 Blake J., Boileau D., Bojunga N., Carninci P., de Bonaldo M.,
 Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
 Gustincich S., Hill D., Hofmann M., Hume D.A., Kanuya M., Lee N.H.,
 Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombauris P.,
 Morone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
 Suzuki H., Toyooka K., Wang K.H., Weitz C., Wittaker C., Wilming L.,
 Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlsaki S.,
 RA Hayashizaki Y.;
 RT *Functional annotation of a full-length mouse cDNA collection.*
 KL Nature 409:685-690(2001).
 DR EMBL: AK018602; BAB31303.1;
 DR MGD: MGI:1918836; 9130009C22RLK.
 SO SEQUENCE 210 AA; 24441 MW; 2EBFD06343933DF CRC64;
 Query Match 18.1%; Score 960; DB 11; Length 210;
 Best Local Similarity 86.7%; Pred. No. 19e-45;
 Matches 182; Conservative 10; Mismatches 18; Indels 0; Gaps 0;
 OY 816 MVQARARADSTVYLVHSGSGVIEHETVNDPREKMYKAIHCVQNMKDEFAVHKLLE 875
 DB 1 MVQARARADSTVYLVHSGSGVIEHETVNDPREKMYKAIHCVQNMKDEFAVHKLLE 60
 OY 876 LQMSIMKMKTKRNIAKHYKKNNSULITFLCKNSVLAQSGEDHIVIEKMIHVMTEPE 935
 DB 61 LQMSIMKMKTKRNIAKHYKKNNSULITFLCKNSVLAQSGEDHIVIEKMIHVMTEPE 120
 OY 936 KELIYIRENKALOKKCAQOINGELIICGQAMGTMMVHKSJDLATKIRNVVVFKNNS 995
 DB 121 KELIYIRENKALOKKCAQOINGELIICGQAMGTMMVHKSJDLATKIRNVVVFKNNS 180
 OY 996 TKQYKKWELPITPNDLYSECLFSD 1025
 DB 181 PKQYKKWELPITPNDLYSECLFSD 210
 RESULT 13
 OY4165 PRELIMINARY: PRT: 1037 AA.
 AC O4165;
 DT 01-JUN-1998 (TREMblrel. 06, Created)
 DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
 DE Hypothetical 119.2 kDa protein.
 GN F15B10.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabdilitida;
 OC Rhabdilitida; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxId=6239;
 RN (1)
 RP STRAIN=BRISTOL N2;
 RC MEDLINE=99069613; PubMed=9851916;
 RA None;
 RT *Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium.*
 RL Science 282:2012-2018(1998).
 RN (2)

RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Mohlmann P., Murray J.;
 RT "The sequence of *C. elegans* cosmid F15B10."
 RN Submitted (Dec-1997) to the EMBL/Genbank/DBJ databases
 RL
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RT "Direct Submission."
 RN Submitted (Aug-2001) to the EMBL/Genbank/DBJ databases
 RL EMBL: AF036696; AAB88350.1;
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR001650; Helicase_C.
 DR Pfam: PF00270; DEAD; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR SMART: SM00487; DECD; 1.
 DR SMART: SM00490; Helic; 1.
 DR ATP-binding; Helicase; Hypothetical protein.
 SQ SEQUENCE 1037 AA; 119188 MW; 63D189175DDA6776 CRC64;

Query Match 12.9%; Score 687; DB 5; Length 037;
 Best Local Similarity 25.0%; Pred. No. 1.7e-29;
 Matches 251; Conservative 169; Mismatches 408; Indels 174; Gaps 35;

OY 101 DLPSPEFAHADE--YLOLLNLOPTLVYDKLLVRODYDKMEEE----- 142
 DB 63 DLKSLSLNADDERLYKIDIMTYLO--LYPKCYVHKLLN--SNREVKLSFHYLDHEFF 120
 OY 143 LTTIEDNRNIAAENNGNSGVRELKTRVQK---ENMFSAFLN--VLK--G--NNELVQ 195
 DB 121 LRFIEKRVAVLYDSYPTIDAVALRKIEERNEEDNODSDFKLKLLRVPVLGGQAVY 180
 OY 196 EL--TSGDSESAELENLSQVDGPOVEQQLSTVQPLKEVWCMEN-----N 243
 DB 181 DLMYITSEKSNLDVE-----AKQFLA-----KYLRLKRLGFLRFYQIIN 221
 OY 244 SSESFASSSVSESDTSLAEGSVCLDSLGHS-----NMGSDSGTMSGDSDE--ENV 296
 DB 222 ASRQNLNGRIYICPVHESATEMMVYIGTALNTNRKRMNINRYDNVQENRTPRLVRSV 281
 OY 297 AARASEPELQLRPYOMEVAOPALEGRNITICLPISGKFRVAVYIAKHLDKKKASRP 356
 DB 282 RQRIHQROLCLRNQOEELQYVALQCKNTIVAPTSGKIVTAANIKHEESRSESKR 341
 OY 357 GKVIYLV--KVLLVQELFKREFQPLKRYVIGLSGDTQLKISFEVYKSCDIIISTQ 415
 DB 342 FRALEFPTPSMILNQ--AASISSYLDHYHTQIIOGSD--NVPANVLKSKDLVAIYO 397
 OY 416 ILENSLLNLENGEDAGVQ-----LSDFSLITIDECHTKKEAVYNNIMRYMOKIKNR 470
 DB 398 MLYNLCNEHRNSLDDSRDLQDFLSTFTITTFDECHNTAKNSPSYIMEH--HYLKNKG 455
 OY 471 LKKENKPVLPOLIGLTASPGVGATKQAKAEHILKLCANDATITIKVENIDOLAN 530
 DB 456 NMPEGH--SLPQIIGLTASTIGTGKNDQVANYIAGCASMDVADLSIAKNDLELHG 512
 OY 531 QIOPECKKFAIADATREDP---FKELLLEPTRIQ-----LHC 565
 DB 513 YSDIVDPKVALCERSTGDFIGMTNRLTLMQVEGLIKALNENHIGIPQRQIETIR 572
 OY 566 QMSPMGDFTOP-----YEOHAIQMEKKAARKGRKEVCAEHLRKKNFALOINDIIR 618
 DB 573 DRPDSSEFLDPPADKEHAGYQNVV-----NOMLVSGTSFRERQ--RTILDEAD 622
 OY 619 MI-DATVHLETFYNEKDKKFAVIEDSDSGDDEYCDDEDEDLKKPKLDETDKFLM 677
 DB 623 VLKECCCTLSYNINFPVALNLKDEM-----EY-----RTNFTIV 659
 OY 678 TLEFENKMKRLAENPEYENKELIKLRNTIMEQYTRTESAGLIPTIKOSAVALSWM 737
 DB 660 NMIRIRERYHNOLVGTGSAENPMSIKTVQYIVQONLORADS--RTILFVRIKEATILNRV 718

OY 738 ITENEKFAEVGKVAHHLIGAGHSEFK---PMTQNEKEVISKPRCTKINLLIAVAEE 794
 DB 719 LNSRELLMLGKSEBMHGLNKSTASSADISAKOKOMEKLMKADGRIIRLVISAEE 778
 OY 795 GLDIKECNIVIRYGLVTEIAIMVQARCARADESTYVLAHSGGVIEHETVNI--REKMM 854
 DB 779 GLDVECSLVIRKYNATNEIAHVORCRGRALNSCVLITNS--IALRQFESNMK--NSLM 837
 OY 855 YKALICVONMKPE-----EYAKHLELOMOSIMEKKKTRNIAKHKNPSSIIFLCK 908
 DB 838 SETISLNSPAEFKRCVDEESNKTPRIREDTDKAKIEQI-----NNIVAKILCK 892
 OY 909 NCSYACSGEDIVIEKHHVNMTPF-----KELYIYENKALQCKADYOINGELICK 963
 DB 893 KCEALICTSKDIR--SRNTOYLCDPFGFSLVKKRTLTDEQALIK-----YATGSLNCR 947
 OY 964 --CGQAMTMMVHKGDLPLCKIRNFVYVFNKNSKQYKRM 1003
 DB 948 ENCGIKLQLELVNTVDPLCLSLALSTIVLVECTDRRIIVKWM 989

RESULT 14
 OY 08VE79 PRELIMINARY; PRT; 143 AA.
 AC 08VE79;
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Similar to RIKEN cDNA 9130009C22 gene (Fragment).
 GN 9130009C22RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 NC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (Dec-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: BC019605; AAB19605.1;
 DR MCD; MGI:1918836; 9130009C22RIK.
 FT NON-TER
 SQ SEQUENCE 143 AA; 16742 MW; 7F16E2CB4090AABA CRC64;

Query Match 12.8%; Score 678; DB 11; Length 143;
 Best Local Similarity 87.4%; Pred. No. 3.5e-30;
 Matches 125; Conservative 6; Mismatches 12; Indels 0; Gaps 0;

OY 883 EKKMKTRNIAKHYNPSLITFLCKNCSYVLAGSGEDIVIEKHHVNMTPFCKLYIVR 942
 DB 1 EKKMKVYKSIAMQYNDNPSLITLLCKNCSMLVCSGENIHVIEKHHVNMTPFCKLYIVR 60
 OY 943 ENKALQCKADYOINGELICKCGQAMGTMMVHKGDLDECLAIRNVVYFNKNSKQYK 1002
 DB 61 ENKALQCKADYOINGELICKCGQAMGTMMVHKGDLDECLAIRNVVYFNKNSKQYK 120
 OY 1003 WVELPTPEPNLYSQCFLPSDED 1025
 DB 121 WVELPTPEPNLYSQCFLPSDED 143

RESULT 15
 ID 017545 PRELIMINARY; PRT; 811 AA.
 AC 017545;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Hypothetical 93.4 kDa protein.
 GN C0B10.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 NC Rhabditidae; Peloderinae; Caenorhabditis.
 NX NCBI_Taxid=6239;

RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2.
 RA MEDLINE=99069613; PubMed=9851916;
 RT None;
 RT "Genome sequence of the nematode *C. elegans*: a platform for
 RT investigating biology. The *C. elegans* Sequencing Consortium.";
 RL Science 282:2012-2018(1998).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2.
 RA Blanchard M., Bradshaw H.;
 RT "The sequence of *C. elegans* cosmid C01B10.";
 RL Submitted (May-1996) to the EMBL/Genbank/DBJ database.
 RN (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2.
 RA Waterston R.;
 RT "Direct Submission.";
 RL Submitted (JUN-2001) to the EMBL/Genbank/DBJ database.
 DR EMBL: U58757; AAD47117.1; .
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR001650; Helicase_C.
 DR Pfam: PF00270; DEAD; 1.
 DR Pfam: PF00271; Helicase_C; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR SMART: SM00490; HELIC; 1.
 KW ATP-binding; Helicase; Hypothetical protein.
 SO SEQUENCE 811 AA; 93435 MW; 7FFAC1860EBBFOA3 RC64;

Query Match 12.3% Score 654; DB 5; Length 811;
 Best local similarity 27.0%; Pred. No. 7.7e-28;

Matches 210; Conservative 147; Mismatches 318; Indels 102; Gaps 27;

QY 294 ENVAAPAPPELOLPYMEVAOPALEGKNIILCLPTSGRIYVAAYIAKDLKKKA 353
 DB 49 ESYRORIHIOFOLRNYOELOVALQCKNTIVTAPGSGKIVIAANIKHEFESRSE 108
 QY 354 SEGGKIVLV-NKVLVLEDFKREPPLKKYRVIGLSDTQIKISPPYVKSCDIIS 412
 DB 109 GKFKALFMTNPMNMLNQGSTRIS-----YLTQTQSD-NVPTNNVIOSKDLIVA 159
 QY 413 TAOILENSLNLNGEDAGVO-----LDSFLIIDECHHINKEAVYNNIMRHLMCK 465
 DB 160 TPQMIVN-ICN-EHRPRLDDEYPPROFLSTIIIFEDYCHNIVKNSPTSNWREY--HY 215
 QY 466 LANNRLKKNKRVILPQILGLASPGVGAATKAAERELIKVANDAFITIKYKENTL 525
 DB 216 LKNGMMPGHH--SFPQIIGLASIGTDKKNCKMGOVRSYIAGCANMDVRELSTVKDNL 272
 QY 526 DOLKNOIOEPCKKFAIADATREDP-----FKELLETMTRIJYCCOMSMDSFGTDP- 577
 DB 273 EELLDNPFVTDQVSCFSCNSDGPLETEFKRLKQMOVEEDLIRTLKNEPTVKYEIPPT 332
 QY 578 -----YEQMAYOMEKKAAGKGRKERVCAEHLRYNEALQINDITRMVATHELETYN 641
 DB 333 DKEHNTEYEMINQNRNCVSLAGRNKTLILEVL-----IVLKVOYHNSOKMIFWN 382
 QY 632 BEKDKKFAV--IEDSDSGDDEYCDGDEDEDLKKPLKID-----LIDRPLMTLFF 681
 DB 383 YFRKRYFGFKRIFOFODCFYALSYNINFPBEVALKKYLEKELDEPERIRNFDN--MNLW 440
 QY 682 ENNKMLKRLAENPVEYENELTKLRNTIMQYTRTESARGLIFKTRQSAVALSQMLTEN 741
 DB 441 DN--CHRELVGIGSAENPMIARTVOITLDONEQTSDFRATLIFVTRKKEADFLANYLV--N 495
 QY 742 EKFAEYGVAAHLIG---AGHSSEFFPMTONOEKVEISKFTQKINLLIAITVAEGID 798
 DB 496 DLHLGLITSDMMSGOKSTASADISAKQOMKELKMFADPNOILLVSTVAEEGID 555
 QY 799 KECNIVIRYGLVTNETIAMVQARGARADESTVYLAHSGSGVLEH--ETVNDPREKMYK 856
 DB 556 PCSLIVIKYVATNETIAHVQRGRARARNSKCVLLTNS---IALHVQESNNLAKENIMTE 612

QY 857 AICHVONMKPE-----EYAHKILELQMSIMEKKMKTKRNIAKHVKNPNSLIFLCKNC 910
 DB 613 TSLIÖNSPGEFRQCVDEESNKKVMPRIQREDDTKAORIKEQI-----NRNIYKIVCKKC 667
 QY 911 SYLACSGEDINHIEKMHVNMTPF-----KELYIVREKKAIOKKCADYQINKEIIG---K 963
 DB 668 DTVLCTNKDIR-SKNIOYIVCNPGFWSLVRRILPDELORASNK-----FNSTGSHIGLGR 722
 QY 964 CGOAMGTMMVHKGDLDPCLIKIRNFYVFEKNNSTKKQYKKWVEL-----PIYPRNLD 1014
 DB 723 CGSKIGQLIDVNTVNLPLCAKVSITILLTESTNERILVQOMKNILDEHPTPTIKKGD 779

Search completed: May 8, 2003, 15:54:11
 Job time : 56 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 15:46:03 : Search time 27 Seconds

(without alignment)
1574.564 Million cell updates/sec

Title: US-09-515-363C-2

Perfect score: 5311
Sequence: 1 MSNGYSTDENFYLISCFA.....LPITFVNDLYSCQLEFDED 1025

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112692

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	402.5	7.6	1374	YC9A_SCHPO	C09884 schizosacch
2	392.5	7.4	778	YF05_MERJA	F58900 methanococ
3	350	6.6	1845	YM68_CAEEL	F45529 caenorhabdi
4	303.5	5.7	993	YIS2_YEAST	F40562 saccharomyc
5	223	4.2	557	YOH8_BACSU	F54509 lactillus su
6	217.5	4.1	2653	CENE_HUMAN	C02224 homo sapien
7	214.5	4.0	1790	USO1_YEAST	F25386 saccharomyc
8	204.5	3.9	2163	BRR2_YEAST	F12639 saccharomyc
9	197.5	3.7	1163	SBCC_GLOAB	C97481 clostridium
10	197	3.7	2230	GOG4_HUMAN	C13439 homo sapien
11	193.5	3.6	663	UVRB_STRAA	C09417 staphylococ
12	193.5	3.6	2017	MYSN_DROME	C49323 drosophilla
13	193	3.6	715	HEIS_PYRAB	C00049 pyrococcus
14	191	3.6	3911	AKA9_HUMAN	C06966 h-a-kinase
15	189.5	3.6	3210	CENE_HUMAN	F49454 homo sapien
16	188.5	3.5	1433	REST_CHICK	C42184 gallus gall
17	187.5	3.5	720	UVRB_PYRPU	C73946 pyrococcus
18	183.5	3.5	663	UVRB_FUSNN	C09192 fusobacteri
19	183.5	3.5	1875	MLP1_YEAST	F12455 saccharomyc
20	182.5	3.4	3660	DMD_CHICK	F11533 gallus gall
21	181.5	3.4	2869	RBP1_PLAVB	C00793 plasmodium
22	180.5	3.4	479	DBPA_BACSU	F42305 bacillus su
23	180.5	3.4	976	SCPI_HUMAN	C15431 homo sapien
24	180.5	3.4	1131	YAB9_YEAST	F81360 saccharomyc
25	180	3.4	2704	BPA1_HUMAN	C03001 homo sapien
26	179.5	3.4	607	DB10_NICSY	F46942 nicotiana s
27	179.5	3.4	1938	MYH6_MOUSE	C02566 mus muscula
28	179.5	3.4	2022	ANTI_ONCVO	F21249 onchocerca
29	179	3.4	564	ROK1_YEAST	F45818 saccharomyc
30	179	3.4	662	UVRB_STEEN	C04966 streptococ
31	179	3.4	754	YAJ3_SCHPO	C00903 schizosacch
32	179	3.4	2748	NUM1_YEAST	C00402 saccharomyc
33	178.5	3.4	1539	Y373_HUMAN	C15078 homo sapien

34	177.5	3.3	444	1	SRMB_PCOLI	P21507 escherichia
35	177	3.3	1679	1	Y109_YEAST	P44572 saccharomyc
36	176.5	3.3	2116	1	MY52_DICDI	P04791 dictyosteli
37	176	3.3	658	1	UVRB_LISTIN	O94783 listeria in
38	174	3.3	687	1	DB73_DROME	P26490 drosophilla
39	173.5	3.3	1227	1	ALM1_SCHPO	O94003 schizosacch
40	173	3.3	637	1	NTP1_FOMPV	O72907 fowlpox vir
41	173	3.3	2349	1	TPR_HUMAN	P12272 homo sapien
42	170	3.2	1102	1	MYSC_CHICK	P20016 gallus gall
43	169	3.2	658	1	UVRB_LISMO	O84415 listeria mo
44	168.5	3.2	646	1	UVRB_METTH	O26542 methanobact
45	168.5	3.2	748	1	CLPE_LACIA	O94103 lactococcus

ALIGNMENTS

RESULT 1	ID	YC9A_SCHPO	STANDARD:	PRT:	1374 AA.
AC	C09884:				
DT	01-FEB-1996 (Rel. 33, Created)				
DT	01-FEB-1996 (Rel. 33, Last sequence update)				
DT	15-JUN-2002 (Rel. 41, Last annotation update)				
DE	Putative helicase C584.10C (EC 3.6.1.-).				
GN	SPCC584.10C.				
OS	Schizosaccharomyces pombe (fission yeast).				
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;				
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;				
OX	NCBI_TaxID=4896;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=972;				
RC	MEDLINE=21848401; PubMed=11859360;				
RA	Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,				
RA	Sgourou J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,				
RA	Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,				
RA	Collins M., Connor R., Cronin A., Davis P., Felwell J., Fraser A.,				
RA	Gentles S., Godle A., Hamlin N., Harris D., Hidalgo J., Hodge G.,				
RA	Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jarvis K.,				
RA	James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,				
RA	Mooney P., Moule S., Mungall K., Murphy L., Nibbellett D., Nisell J.,				
RA	Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch F.,				
RA	Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,				
RA	Skellton J., Simmonds M., Squares R., Squares S., Stevens K.,				
RA	Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,				
RA	Woodward J., Volckaert G., Aert R., Robben J., Grimprey P.,				
RA	Welffens I., Vanstreels E., Rieger M., Schaefer M., Muelier A.,				
RA	Cabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hillert H.,				
RA	Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl J.M.,				
RA	Eger P., Zimmermann W., Medler H., Wandut R., Purnelle P.,				
RA	Galbault A., Gadieu E., Dreano S., Gloux S., Lelure V., Muller S.,				
RA	Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,				
RA	Lucas M., Rochet M., Gallardin C., Tallada V.A., Garcon A., Jude G.,				
RA	Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forstner S.L.,				
RA	Cerrutti L., Lowe F., McCombie W.R., Paulsen I., Potashkin I.,				
RA	Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.,				
RT	The genome sequence of Schizosaccharomyces pombe.*;				
RL	Nature 415:871-880(2002).				
CC	-1- SIMILARITY: BELONGS TO THE HELICASE FAMILY.				
CC	-1- SIMILARITY: CONTAINS 2 RNASE III DOMAINS.				
CC	-1- SIMILARITY: CONTAINS 2 RNASE III DOMAINS.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL database				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	EMBL: AL032824; CAB37423.1; -				

DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR005034; DUF283.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR000999; RNase_3.
 DR Pfam: PF00270; DEAD_1.
 DR Pfam: PF00271; Helicase_C_1.
 DR Pfam: PF00636; Ribonuclease_3_2.
 DR Pfam: PF03368; DUF283_1.
 DR SMART: SM00490; HELIC_C_1.
 DR SMART: SM00535; RIBOC_2.
 DR PROSITE: PS00517; RNase_3_1; 1.
 DR PROSITE: PS0142; RNase_3_2; 2.
 DR Hypothetical protein; Helicase; ATP-binding; Hydrolase; Nuclease;
 KW Endonuclease; Repeat.
 FT NP_BIND 32 ATP (POTENTIAL).
 FT SITE 145 DEH BOX.
 FT DOMAIN 916 1038 RNase III 1.
 FT DOMAIN 1083 1233 RNase III 2.
 SQ SEQUENCE 1374 AA; 158039 MW; 89AE9EFDE7966C6 CRC64;

Query Match

Best Local Similarity 7.6%; Score 402.5; DB 1; Length 1374.
 Matches 175; Conservative 107; Mismatches 244; Indels 157; Gaps 32;

QY 304 PELQRPYQHEVAPALEGKNIITCLPTSGKTRAVAYIAKHLDKK-----KKA 353
 9 POL-LRKYQODVYNIASK-ONTLLYMTGAGKTLIAVKLIKOLEOILLOESNLEHKKI 66
 DB 67 S-----VELNKVPLVFOQAEYIRISOLPAKVMFYGELSIEMSEQL--INILRYNVI 118
 QY 354 SEQKQIVLVNKKLVLEO---LPRKEFOPFLKKWRYVIGLSDTOLKISPEVYKSDII 410
 DB 67 S-----VELNKVPLVFOQAEYIRISOLPAKVMFYGELSIEMSEQL--INILRYNVI 118
 QY 411 ISTAQLIENSLNLENGEDAGYOLSDSLIIDECHEHTNKEAYNNIMKRYLMOKLKNK 470
 DB 119 VITADLF---YLTLARG---FLSINDMLIIECHALINDAYARIIMLY-----HR 166
 QY 471 LKKE-NKVPILPQILGLTASPGVGTAKQAKAEHILKICANDIAFTIKVENLQIK 529
 DB 167 AKAVLSKKHFTLPRIFGMTASP-----FTCK--KGNLYHRI 200
 QY 530 NOIOEPCPKFAIDATREDPFKEKLEIMRTIYTCOMSPISGFGTQPEYQMAIQEKKA 589
 DB 201 YQ-----WEOLFDSKAV---VSENILADYFGLPESYVYMSNKL 218
 QY 590 AKGNKREKRYCAEHL-----RKYNDALOINDTIM-----IDATHELEFENEKIKK 637
 DB 239 VPPSDSIKKCEETLOGCKLISAVKATA--ETIDMGLMFGCVWLIVLDEVETRLKK 296
 QY 638 FAVIEDSDSGDDEYCDGDEDEDLKKPLKIDETRFIMTFFENNMKIKRLAEN-EYE 697
 DB 297 KALGKQLSD-----DEE-----LAIDRLKIEVED--WKNNKYSINGRIPVFD 337
 QY 698 N-----EKILKRLNTIMEQYTRTESARGIIFETRSATAALSWITENKFAEVGYKANH 753
 DB 338 STDYTKVKKVLELLKATY-RKSDSVRTYIPEKKAFAETLSLFM---KILMLPNIKAS 393
 QY 754 LIGAGHS--SEKPMTONBOKEVYSKFTGKIILATVAEAGLIDIKENIVIRIGLT 811
 DB 394 FIGHGSPDGGES-KTFPRQKOTLHKFKTGKYNVLATVAEAGIDVPSNLTIRINIKR 452
 QY 812 NEIATYQANGARADESTYLVLAHSGGYIE---HEIVN-----DEFTKMKYKAHCV 861
 DB 453 TVTQVYQSKGRARAKASKFLIPNTBELIHERILHEEKNLFAELSELNSNIFOSLVE 512
 QY 862 QNNK-PEEYAAKLE-----LQNSIMEKKMKT-KRNIAHY-----KNNSDLITLC- 907
 DB 513 EREKVTDDIVYEGETGALLTGIVAVSLYNFCNTLSROYTRYVPTFTADQLSGWCF 572
 QY 908 -----KNCYVLACSGEDIVHYEK 925
 DB 573 EVELPRAKCYVPAAGSPASIKK 595

RESULT 2
 YP05_MET2
 ID YP05_MET2A STANDARD; PRT; 778 AA.
 AC Q58900;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Putative ATP-dependent RNA helicase M01505.
 GN M01505.
 OS Methanococcus jannaschii.
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
 OC Methanocaldococcaceae; Methanocaldococcus.
 OX NCBI_TaxID=2190;
 RN 11;
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 Sutton G.G., Blake J.A., Fitzcerald L.M., Clayton R.A., Gocayne J.D.,
 Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Belch S.L.,
 Overbeek R., Kirkness E.F., Weissflog K.G., Merrick J.M., Glodek A.,
 Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
 Ulteback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RA Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.
 RT Science 273:1058-1073(1996).
 RL -1- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY. DEAD
 SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

CC EMBL: U67591; AAB99518.1;
 CC DR TIGR: M01505;
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002464; DEAD_box.
 DR InterPro: IPR004016; ERCC4.
 DR InterPro: IPR003583; HHH_1.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR00445; HHH.
 DR Pfam: PF00270; DEAD_1.
 DR Pfam: PF00271; Helicase_C_1.
 DR Pfam: PF00633; HHH_1.
 DR Pfam: PF02732; ERCC4_1.
 DR SMART: SM00487; DEHC_1.
 DR SMART: SM00490; HELIC_C_1.
 DR SMART: SM00278; Hhh1_1.
 DR PROSITE: PS00690; DEAD ATP HELICASE; FALSE NEG.
 KM Hypothetical protein; ATP-binding; RNA-binding; Helicase;
 FT NP_BIND 35 ATP (POTENTIAL).
 FT SITE 137 140 DEAD BOX.
 FT DOMAIN 532 535 POLY-GLU.
 SQ SEQUENCE 778 AA; 88882 MW; DCB1F220801338D8 CRC64;

Query Match 7.4%; Score 392.5; DB 1; Length 778;
 Best Local Similarity 22.9%; Pred. No. 8.5e-13;
 Matches 169; Conservative 128; Mismatches 235; Indels 205; Gaps 28;

QY 306 LQRLPYQHEVAPALEGKNIITCLPTSGKTRAVAYIAKHLDKKKASERKXIVIVNK 365
 DB 13 LEARLYQOIIANNAALKKTKLVC-ISTGLKTAIALVLVAGILTK-----DKKVIILAPS 66
 QY 366 VLVLEOLPRKEFOPFLKKWRYVIGLSDTOLKISPEVYKSDIIISTAQIIEHSLNLE 425
 DB 67 RPLVQHNHNRILKQVNLINDEDKIATLTKIQPK-KRAELYKKGKIFIAIPQVINDIIT 122

```

QY 426 NGEADG-VOLSDPSLLIIDECHHTNKEAVYNNIMRHYLMOKLKNRLKK: HKPVILPOT 484
DB 123 -----AGRINDEFILLADRAHHTTGDHAY-----AFVAKKFKOK-----CHT 161
QY 485 LGLTASPGVGATKQAKAEHILKLCANLDAFTIKTVKINLDOCK-----529
DB 162 LGLTASPG-----SDIDKVMCEINIGIEHVEVTRHEDVOKPYIAVKILPILRID: 213
QY 530 -NOIOECKFALADTRE-----DPPKLEIMRIQVCOMKWSOP 573
DB 214 PNEFKALK--LINEALKERKLKLDAGVINSTADYKTELIDNKNL:-----262
QY 574 GTOPYEOMATOMEKKAAGNKRERKVCARLRYNEALONDIIRMDA:-----1THUD 627
DB 263 -----DEEVYELIKYCEALKIMAKELLES:OKSVFVINYIN 300
QY 628 TFYNEKKKKFAVIEDSDGDEYCDGDEDEDLKKPLKDETRFLMLFFENNMKL 687
DB 301 KLSMORTKSAKSIVNDE-----KVRERAVNLMKKSQVHEPKL: 336
QY 688 KRLAENPEYENKLTIRNTIMEQYTRTESAGIIFTKROSAYALSON: TENEKFAEY 747
DB 337 -----GKYVDVKNKILEK-----NKDERIIIFADYRDVKEKIVNI:TON-----375
QY 748 GYKAHHLIGAGHSSEFPKPTONOEKEYISFKRTGKILNLATIVAEGL: KECNIVIRY 807
DB 376 GYAIRITGGA-NKEGKMSQKEQIEAIEFKK-EGSVIVTSVSEBGI: PBYVNIIFV 433
QY 808 GLVTNELIAVQANGRAADE-STYVLVAHSGSGVIEHDI:VNDPERKMA:NAIHCVQNM- 864
DB 434 EYPVSEIRFIORGRARCEGKVVYLAKGTADAEAYRSAL:YKEREMKIL:---KNMG 489
QY 865 -----KPEYAKHILELOMSITMEKMKRKNRINAKIKY:-----NINSLITFLCKN 909
DB 490 YLNLKRLQKFEESKEKEETEKEIEKEIEKESKTAKERTKEEEKTKKIVTILDFIKO 549
QY 910 CSV--LACSGED-----IHVIEKHHVNMTP:-----EFK-1:-----YVRE 943
DB 550 IEVKEKSEKEDKIKOPIKIPKRIKITIVYREKMKALHANYANIELKILEVGDVYLSND 609
QY 944 NKALOKKADYQINGEI 960
DB 610 RYVVERKTADEPVNSII 626

```

RESULT 3

ID	YMG8_CAEEL	STANDARD	PRT	1845 AA
AC	P34529			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Hypothetical helicas K12H4.8 in chromosome III.			
GN	K12H4.8			
OS	Caenorhabditis elegans.			
OC	Eukaryota: Metazoa: Nematoda: Chromadorea: Rhabditida: Rhabditidae:			
OC	Rhabditidae: Petodermidae: Caenorhabditis.			
OX	NCBI_TaxID=6239.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Bristol N2.			
RX	MEDLINE=94150718; PubMed=7906398:			
RA	Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,			
RA	Craxfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,			
RA	Bonfield J., Dear S., Du Z., Durbin R., Favello A., Fraser A.,			
RA	Fullston L., Gardner A., Green P., Hawkins T., Hillier L., Jiet M.,			
RA	Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,			
RA	Lattelle P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,			
RA	Parsons J., Percy C., Rifen L., Roopra A., Saunders D., Shownkeen R.,			
RA	Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,			
RA	Stulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vautian K.,			
RA	Waterston R., Watson A., Weinstein L., Wilkinson-Sproat J.,			

```

RA Wohldman P.:
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.
RL Nature 368:32-38(1994).
RN (2)
RP REVISIONS.
RA Waterston R.:
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE HELICASE FAMILY.
CC -1- SIMILARITY: CONTAINS 2 RNASE III DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 PAZ DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 DRBM (DOUBLE-STRANDED RNA-BINDING) DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.sib.ch).
CC
CC EMBL: L14331; AAA28101.2; -.
CC PIR: S44849; S44849.
CC WormRep: K12H4.8; CE25057.
CC InterPro: IPR001410; DEAD.
CC InterPro: IPR001159; DS_RBD.
CC InterPro: IPR005034; DUF283.
CC InterPro: IPR001650; Helicase_C.
CC InterPro: IPR003100; PAZ.
CC InterPro: IPR000999; RNase_3.
CC Pfam: PF00270; DEAD_1.
CC Pfam: PF00035; dsrm_1.
CC Pfam: PF00271; helicase_C_1.
CC Pfam: PF00636; Ribonuclease_3_2.
CC Pfam: PF02170; PAZ_1.
CC Pfam: PF03368; DUF283_1.
CC SMART: SM00487; DEXDC_1.
CC SMART: SM00358; DSRM_1.
CC SMART: SM00490; HELIC_1.
CC SMART: SM00535; RIBOC_2.
CC PROSITE: PS50137; DS_RBD_1.
CC PROSITE: PS50821; PAZ_1.
CC PROSITE: PS00517; RNASE_3_1; 1.
CC PROSITE: PS50142; RNASE_3_2; 2.
CC Hypothetical protein; Helicase; ATP-binding; Hydrolase; Nuclease;
CC Endonuclease; Repeat.
CC NP_BIND 33 40
CC FT DOMAIN 759 915
CC FT SITE 145 148
CC FT DOMAIN 1316 1524
CC FT DOMAIN 1578 1740
CC FT DOMAIN 1768 1831
CC FT DRBM.
CC SQ SEQUENCE 1845 AA; 210922 MW; 4A96EA2922FF1D9A CRC64;

```

Query Match

Best local similarity 23.5%; Score 350; DB 1; Length 1845.

Matches 175; Conservative 118; Mismatches 249; Indels 2; Gaps 37;

```

QY 309 RPYOMEVAOPALEGKNIICLPYSGKTRVAVYIANDH-----LDKRRKASE:NAVIV 361
DB 14 RDYQVELDKATK-KNTIVOLGTGSGKTFIAVLLEKEYGVQLFAPLDQCK:---KAFV 67
QY 362 LVVKKVILVQLFKEKPOPLKKRYVIGLSGP-----QKISFPEVAVSCIIISAD 416
DB 68 VKEVKNVLEQ--QAIHIEVHNSFKVGYNIGOTSGLMSKECCDFMKRHIVVITAC 124
QY 417 LENSILNLENGEDAGVLSLIIIDECHHT-NKEAVYNNIMRHYLMOKLKNRLKKN 475
DB 125 L-----LDLIRAVYLIKEDKCYLIPDECHNALGSHPRKSIKVDY-----KIKMD 170
QY 476 KVPVLPOLIGLTPASPGVGATKQAKAEHILKLCANLDAFTIKTVKINLDOCK:IDEP 535
DB 171 KPV---PVLVGLTASL-IAKAVAPERKLMQKLKLESAMD-S-VLETSND-LVNS:SVIGAMP 224

```

OY 536 -----OKKFAIA--DAREDPREKLELEIMTRIQTQWSPMSNDSTQ-----576
 DB 225 YEAVIICKDEIGCGIPNEDYIEIDEVAVANITTEPHPIDDEPRHFKDSEKIFR 284
 OY 577 -----PYEOMAIOMKKAOKKONKREYCAELHLYN:ALOJND-T:LNMIID-AV:HLI: 627
 DB 285 AVFROLOPNAAM-----R:AOVWEKELGKTIKSQV:POKTLTFLNMAKISMT 331
 OY 628 TFFNEKDKKFAVIEDDSDEGDEYCDGDEDEDLKKPLKLEIDREFLMTFFENKML 687
 DB 332 T-----KRL-----EPKMKIKSIALPRYVIGVILRLHEIL 365
 OY 688 KRLAENPEYENKLTIKRNTINQYTTESAGIITFK:ROSAVAL-----SCVITE 740
 DB 366 ETF--NPEFOKERN-KLEKA-----EHL:SAIIFVD:KRYAVALSLLMRH:KSW--- 410
 OY 741 NEKFAEYVAAHLIGAG-----HSSEFKPTONEOKKEVISKFRCTGNILATTVAEGL 796
 DB 411 EPNFK--FVNPDYVYGASGNILASSDQGL-HKROFELVLRHFRHNEINCL:ATSVLEEGV 467
 OY 797 DIKCNIVIRYGLVTNEIAVVOAGRARADESTVYLVANSG-----SGVIE 842
 DB 468 DVKOCNIVIKFDRPLDMRSYVQSKGRARRAGSRVILVEHKOTAAVCSKIFSDIFRLVP 527
 OY 843 HETVNDPREKMKYKAIHCVONMKP--EYAHKILELOMOS:IMKKKKKTKNIAKHKNMP 900
 DB 528 HNOIPIEENGVTK--YCAELLPINSPIKHA1-----VLKNPMPNK-----569
 OY 901 SLITFLCKNCSVLASGE-DIHVIEK-----MHVNNTPE-----FKR 937
 DB 570 AQMAVVALEACROGLHEBEDLNDLPGRESIAKLEHIDEPDEYAPGLAAKVGSKKRO 629
 OY 938 LYIVRENKAL-----OKKCADY 954
 DB 630 LYDKIARALNSESFEADKECFY 653
 RESULT 4
 Y152_YEAST
 ID Y152_YEAST STANDARD; PRT; 993 AA.
 AC P40562;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Putative ATP-dependent RNA helicase Y1R002C.
 GN Y1R002C OR Y1R2C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI:TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288c / AB972;
 RA Barrett B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
 RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
 RA Gentsch S., Hamlyn N., Hornslett T.S., Hunt S., Jagsels K., Jones M.,
 RA Louis E., Lye G., Moulie S., Moulie T., Odell C., Pearson D.,
 RA Rajadream M.A., Riles L., Rowley N., Skelton J., Smith V.,
 RA Walsh S.V., Whitehead S.;
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 RM [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288c;
 RX MEDLINE=95282515; PubMed=7762303;
 RA Voss H., Tamames J., Teodoro C., Valencia A., Sengen C., Wleemann S.,
 RA Schwyger C., Zimmermann J., Sander C., Ansoyde W.;
 RT Nucleotide sequence and analysis of the centromeric region of yeast
 RT chromosome IX.
 RL Yeast 11:61-78(1995).
 CC -1-SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY. DEAD
 CC SUBFAMILY.
 CC

CC This SWISS-PROT entry is copyright. It is produced through collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL contribution
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb.ch/announce/>
 CC or send an email to license@isb.ch).
 CC
 DR EMBL: Z38062; CAAB6204.1;
 DR EMBL: X79743; NOT_ANNOTATED_CDS.
 DR PIR: S48436; S48436.
 DR SGD: S0001441; Y1R002C.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR002464; DEAD_Box.
 DR InterPro: IPR001650; Helicase_C.
 DR Pfam: PF00270; DEAD_1.
 DR Pfam: PF00271; Helicase_C_1.
 DR SMART: SM00487; DEXDC_1.
 DR SMART: SM00490; HELIC_1.
 DR PROSITE: PS00690; DEAD_ATP_HELICASE; PAUSE_NEG.
 DR Hypothetical protein; ATP-binding; RNA-binding; Helicase.
 KW NP_BIND 107 114 ATP (POTENTIAL).
 FT SITE 209 212 DEAD_BOX.
 SQ SEQUENCE 993 AA: 114057 MW: 47400950.543171F CRC64:
 Query Match 5.7%; Score 303.5; DB 1; Length 993;
 Best Local Similarity 20.3%; Pred. No. 3.8e-08;
 Matches 143; Conservative 118; Mismatches 246; Indels 14; Gaps 25;
 OY 307 QLRPYQMEVAPALAEKNIILGLPGSGCTRAVAVIADHLDKKKKSPGVIVIVKV 366
 DB 86 EVADQYITVHKS-LFONLCALPTGCKTFIASTVMLVFFRWTKA---KILTAIFR 140
 OY 367 LIVEDFRKEQFQFLKKWRYIGLSG-----TOLKISFEYVASCIIISFAIILNS 420
 DB 141 PLVAGQ-----IKACGIGTIPSDQTAIILDKSRKKRETIWANKRVFAFPYVEND 192
 OY 421 LNLKNGEDAGVQLSDFSIIITDECHHTNKEAVYNNIMHYIMOKLKNKRLKIKPPVP 480
 DB 193 L-----KRGVDPKVDIVGLVDEAHKATGSSAYTNVVF-----IIRFNSSV-- 234
 OY 481 LPQILGLTASPGVCGATKQAKAEHILKAMLDAPFTR-TYKENMDLQKNLOPCKPF 539
 DB 235 --RLALATITP-----ASDEGVQEVYNNMLDISKIEIRFEESMDIVKYKKKKKKKI 284
 OY 540 A1ADATREDPFEKLEIMTRIQTQWSPMSDFTQYEWALOM-----EKAA 590
 DB 285 EV-----PLLEIEDLIE-----QIGMAVKVPLQQAIEIGIYEIRTPGQINAF 327
 OY 591 KKGKREKCA-----EHLRKNEAL-----QINDTIMDAVTHLETFYNEFKKRAV 640
 DB 328 KAMQOSQKLIANPTIPEGIKWNPFIQLLNNGVQMLKRLKIY-GIRTFNFKPK--- 382
 OY 641 IEDDSDEGDDDEYCDGDEDEDLKKPLKLEIDREFLMTFFEN-NKMLKRIALN--FE 695
 DB 383 -----CTEFTTKYMLK-----STNKIAAEFTYHPLKNIKKNQYANYSLEK 424
 OY 696 YENE-KLKLKNTIMEQYTRTESAGIITFKROSAYALSOMI-----738
 DB 425 FVGHCKLQCVDELMDPQKRGSDSRVILFELRSALIEYFIDSVADDOIRPHIIGQ 484
 OY 739 -----TENKFAEYG-----VAAHILJGAGH 759
 DB 485 ARAKGFDEVKYTRKHAPKRRKVERLHRQDEKFLAEKTRANDKLER-SAKRTSSSE 544
 OY 760 SEEFPMPTONEKEVISKFRCTGNILATTVAEGLKECNVIRYGLVINFIAMVQA 819
 DB 545 EAOJGMMQMKKEVIHNFKKGEYVAVLCTSIGEGDLIGEDVLITCTDTSSHKNKVR 604
 OY 820 RGR-ARADESTVYLVANSGSGVIEHETVNDPREKMKYKAIHCVONMK-----FF- 867
 DB 605 MGRTRKRRGKIVLFFSSNESYKFERAMEDYSTIALISKQITQKKSDRIIPHIILHFG 664

0Y 868 -----EYAKHLEIEMLOQSIEMKEKPKRPIAK 894
 Db 665 HELTITINDENIEMEDVEDEVIRATOCMAKKVKPKRAITH 708
 RESULT 5
 YOH_H_BACSU
 ID YOH_H_BACSU STANDARD: PRI: 557 AA.
 AC p54509,
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical helicase yqhH.
 GN YOH_H
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus
 OX NCBI_TaxID:1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168 / JH642;
 RX MEDLINE-97124195; PubMed-8969508;
 RA Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato I., Takeuchi M.,
 RA Kobayashi Y.;
 RT "Systematic sequencing of the 283 kb 210 genes-232 d genes region of
 RT the Bacillus subtilis genome containing the skin element and many
 RT Microbiology 142:3103-3111(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RX MEDLINE-98044033; PubMed-9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloul G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolojin A., Borchert S.,
 RA Borris R., Boutsier L., Brans A., Braun M., Brigneau C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano N., Cettler N.M.,
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehlich S.D., Emerson P.T.,
 RA Enlian K.D., Errington J., Faoret C., Ferrari E., Foulard D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Gellera N.,
 RA Gilm S.Y., Glaser P., Goffeau A., Golligly E.J., Granot G.,
 RA Giuseppe G., Guy B.J., Haga K., Hahsch J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holstappel S., Hosono S., Hult M.F., Itaya T., Jones L.,
 RA Joris B., Katamata D., Kasahara Y., Klatier-Blanchard M., Klein C.,
 RA Kobayashi Y., Koehler P., Koningsstein G., Krogh S., Kuwano M.,
 RA Kunita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medicine C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porvillik S., Prescott A.M.,
 RA Priessein E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Roche A., Roche B., Rose M., Satake Y.,
 RA Sato T., Scanlan E., Seclitch S., Schroeter R., Scofield F.,
 RA Selvauchi J., Sekowska A., Serró S.J., Serró P., Shin H.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tanakoshi A., Tanaka T., Terspila P., Tetsuchi A.,
 RA Totsu V., Uchiyama S., Vandenbol M., Vannier F., Vassallo A.,
 RA Viari A., Wambut R., Wedler E., Wedler H., Wellenreger T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yamamoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumschein E., Yoshikawa H., Yanchin A.,
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis".
 RL Nature 390:249-256(1997).
 RL -1 SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.jsb-slb.ch/announce/>
 CC or send an email to license@slb.ch).
 CC -----
 DB EMBL: D84432; BAI12545.1; -

[illegible]

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Hominidae; Homo;
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93024922; PubMed=1406971;
 RA Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.,
 RT "CENP-E is a putative kinetochore motor that accumulates just before
 RL mitosis.";
 RL Nature 359:536-539(1992).
 RN [2]
 RP CHARACTERIZATION.
 RX MEDLINE=95196755; PubMed=7889940;
 RA Throner D.A., Jordan M.A., Schaar B.T., Yen T.J., Mills N.L.,
 RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
 RL microtubule motor.";
 RL EMBO J. 14:918-926(1995).
 RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE=98437347; PubMed=9763420;
 RA Chan G.K.T., Schaar B.T., Yen T.J.,
 RT "Characterization of the kinetochore binding domain of CENP-E reveals
 RL interactions with the kinetochore proteins CENP-F and HUBB1.";
 RL J. Cell Biol. 143:49-63(1998).
 CC -1- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
 CC KINETOCORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
 CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
 CC AND/OR SPINDLE ELONGATION.
 CC -1- SUBUNIT: INTERACTS WITH CENP-F AND HUBB1 KINASE.
 CC -1- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCORES DURING
 CC CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE. AND IS
 CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
 CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@isb-sdb.ch).
 CC
 DR EMBL: Z15005; CAA78727.1;
 DR PIR: S28261; S28261.
 DR HSSP: P17119; 3KAR.
 DR GeneW: HGNC:1856; CENPE.
 DR MIM: 117143;
 DR InterPro: IPR001752; kinesin_motor.
 DR Pfam: PF00225; kinesin_1.
 DR PRINTS: PRO0380; KINESINHEAVY.
 DR SMART: SM00129; KISC_1.
 DR SMART: PS00411; KINESIN_MOTOR_DOMAIN1; 1.
 DR PROSITE: PS50667; KINESIN_MOTOR_DOMAIN2; 1.
 DR PROSITE: PS50667; KINESIN_MOTOR_DOMAIN2; 1.
 KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
 KW Cell cycle; Centromere.
 FT DOMAIN 1 335 KINESIN-MOTOR.
 FT DOMAIN 336 2471 COILED COIL (POTENTIAL).
 FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).
 FT MP_BIND 86 93 ATP (BY SIMILARITY).
 SO SEQUENCE 2663 AA; 231087 MW; CEFCL3880C8V8C8H8 CK564;
 Query Match 4.1%; Score 217.5; DR 1; Length 2663;
 Best local Similarity 18.6%; Pred. No. 0.007%;
 Matches 209; Conservative 185; Mismatches 421; Indels 307; Gaps 47;
 QY 9 ENFRYLSCFRARVKKYIOVEPVLDYLFPAEKKIOJYVATSGNMVAVELLSTLK 68
 DB 1040 EQGRKIFSLQENKELQOMLESYIAEKQDKTKDKENIKMTLE---NQRHKLGLGFLKK 1096
 QY 69 GVMHLCMTREFV-----EALRTGSPPLAARVNNPDLTDPSFENAHDTYQLNLMLQP 123
 DB 1097 -----QOEIVAOEKNAIKKEGELSRICDRLAEVEEKIKKSSQDLQKQGLLVNCE 1149

QY 124 TLVDKLLVNDV--LDKCMEEELITIE--DRNRIAAEN--NGNDSGVRELK-RIVYNT-- 175
 DB 1150 MSRMOKKINEIENIKENKELNKKELTLEHMETERLELAOKINENYEFVKYIKKYLK 1209
 QY 176 -----NMESAFNLVLRGTG-----NNELVQELTGSCSSNNKTNL 212
 DB 1210 KSETERDHLRGVIRIETENTGLQTEELKIANHILKEHETIDELKRS-VSKIKVQIINF 1268
 QY 213 SOVDGPOV-----EEDLLSTYVQNLKEKYMENSSSEFSFSSVSSNLS 261
 DB 1269 QDLKSHTKLOEIRIVLHDEQELL-----PVRKVSSETGETNNELIELTQSTIKS-IT 1322
 QY 262 LAGSVCSDSESGHNSNGSGTSGSDSDENNAARSPPELOLRVQYKVAQVATE 321
 DB 1323 LARTEM-----ERLRNKEF-----QESQELKSLTKERDNKTIKEALVYKHDVLT- 1368
 QY 322 GKNIILCPTGSGTRVAVYIAK-----DHLDKKAKASEPKGVYLVAKVILVYEL 372
 DB 1369 -----KEHIRETLAKIQESOSKQOSLMMKKEDNETKI-----VPRMYS- 1408
 QY 373 FRKEPPLKKMYRVLIGSGDTQAKISPE---VYKSCDIISTAOILENSILNFMN-- 427
 DB 1409 FPKQDALLRIETEMGLS--KRLQSDHEKMSVAKKEDDLQRLQEVLOSSENG;KENIK 1466
 QY 428 EDAGVQLSPESLIIDECHHTKKEAVYNNIMRHYLMOKIKNNRLKKENKRPVPIVYIGL 487
 DB 1467 EIVAKHLETEEBELKVAHCCLEKDEETIN-----ELRVMLSEKETE----- 1506
 QY 488 TASPGVGATKOAKA-----EEHILKCANLDAFTIKYKENMDLQKQICVYKPF 539
 DB 1507 -----ISTIOKOLEAINDLQKIOEIVKEKQOLNIKOISEVQEVNNEK-GPKILKK-- 1558
 QY 540 AIADATREDPFEKLEIMTRIOTYCO--MSPMSDGTQPYQWAIOMIKKAAKKNR-- 595
 DB 1559 --AKDSALQSIKSMELTNRLQESQEEIOIKKEEKRPQVQALQIEDQCKNKTKKEI 1616
 QY 596 -----KERVQ-AEHLRYKNEALQIN-----DTIRVITVAYT 624
 DB 1617 VAKMSEQEKYQFLKTAVNTEOEMCEIELEKQFQETKINDENITENIKIQLIHE 1676
 QY 625 HLETFYNEKKRKFVIEDSDGDDDECDDEDDIKKPLKLD----- 670
 DB 1677 NLEEMSVYKRR-----DLARSVELTKVERDQKLENIRITTDLEKQELKVIHHL 1730
 QY 671 ----ETDRFLMLTFEENKMLKRLAENPEYENKL-----TKLNTIMVYVTE 716
 DB 1731 KEHOETIDKLRGIVSEKTEINISNMOKDLEHSDALKADQKIOELRIAHNH KQVDFIT 1790
 QY 717 ESARGIIFTKROSVAVALSQWITTEN-----EKFAEVGVANHILGAGHS-----SFFK 764
 DB 1791 DKLRGIVSEKTKLSNMOKOLENSNAKLOEKIOELKANEHOLITLKKOVNETQKVVSTPE 1850
 QY 765 PMTQ--NEQKEVYSFRFGKINILATTVAAEGDLKECNIVIRGLVTNEIAVQVAGGR 822
 DB 1851 QLKQIKQDLSLTKLEIENLNL--AQELHNLDEMK-----SVAKKIKLDN 1893
 QY 823 ARADESTYLVLAHSSGVI-----EH-BTVNDFREKMYKATH 859
 DB 1894 LRRVEETIKLERDQKESLOETKARDLEIOELKTARMLSKHEKTEYDKLRKIKELIQ 1953
 QY 860 GVDNNK-----PEEVAHKLLELOMOSIM-----EKKMKTKKRIAHYNNNSLI 903
 DB 1954 ISDIOKDLKSKDELQKQIOELQKKELOLLRVKEDVNNHSHKINEMEOKQKQFLEN --- 2009
 QY 904 TFLCKNCSVLGSGEDIVIEKMHVNTTPPEKELYIARENK 945
 DB 2010 -YLCK-----CEMDNPOLTKLHE-----SLEIRIVAKER 2039
 RESULT 7
 USOL_YEAST ID USOL_YEAST STANDARD: PRT: 1790 AA.
 AC P25386;

DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Intracellular protein transport protein US01.
 GN US01 OR INT1 OR YD050W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 NX NCBI_TaxID:4932.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-X2180-1A;
 RX MEDLINE-91185402; PubMed-2010462;
 RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
 RA Yamasaki M.;
 RT "A cytoskeleton-related gene, us01, is required for intracellular
 RT protein transport in Saccharomyces cerevisiae.";
 RL J. Cell Biol. 113:245-260(1991).
 RL [2]
 RP SEQUENCE OF 782-1790 FROM N.A.
 RA Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
 RA Kendrick K.E.;
 RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases
 RP [3]
 RP SEQUENCE OF 1-8 FROM N.A.
 RA Bai Y., Symington L.S.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ database.
 CC -1- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
 CC COMPLEX.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
 CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
 CC ER AND THE GOLGI COMPLEX.
 CC -1- DOMAIN: THE POOLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE. COMPOSED
 CC OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
 CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
 CC -1- SIMILARITY: BELONGS TO THE VDP/US01/YEL047C FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: X54378; CA38253.1; -;
 DR EMBL: L03188; AAB00143.1; -;
 DR EMBL: U53668; AAB6659.1; -;
 DR PIR: A38455; A38455.
 DR SGD: S0002216; US01.
 DR InterPro: IPR002017; Spectin.
 DR Transport: Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
 KM
 FT DOMAIN 1 724
 FT DOMAIN 725 1790
 FT DOMAIN 465 487
 FT DOMAIN 991 1790
 FT DOMAIN 1172 1786
 FT CONFLICT 847 847
 FT CONFLICT 924 924
 FT CONFLICT 1253 1253
 FT CONFLICT 1319 1319
 FT CONFLICT 1461 1461
 FT CONFLICT 1581 1581
 FT CONFLICT 1600 1600
 FT CONFLICT 1661 1661
 FT CONFLICT 1772 1772
 FT CONFLICT 1772 1772
 SO SEQUENCE 1790 AA; 206424 MW; 6CE2B215E9F4818 CMC64;
 Query Match 4.08; Score 214.5; Dn 1; Length 1790;
 Best Local Similarity 19.38; Pred. No. 0.0027;
 Matches 180; Conservative 158; Mismatches 336; Indels 259; Gaps 37;
 38 LPAAVEQDQRTVATSGNQAVELLSTLEKGVHNLQMTAEFVALRRTS----- 88

Db 904 LKEDIAAKITEIKAINENLEEMKIQCNLSKEEHI--SKELVEYKSRFSOHNLVAKLT 961
 QY 89 -----PLAARY--ANPELTOLPSPFENADEVYQLQNLNLOPLTVLQKL-----LVREV 134
 Db 962 EIKISLANNKKDQAENESIKAVESKNSSYQLSNL--QNKIDMSQEKINPQI)FGS 1019
 QY 135 LDKMEELLITIDRRIRIAAENNGNSGVELIKRIYQKEMFSAFLVLRQ-----T 188
 Db 1020 IENITDQKTTISDL-----EOTKREIISKSDSKDEYSQISLKKLELAT 1068
 QY 189 GNNELVQELTSGDSESNMAEINL)SOVDSPOVEQLLSTTVQNLKEYWGMENNSS 248
 Db 1069 ANEENNKI--SLTTRTELE-----AELAAKYNLKNEL-----ETRLITS 1108
 QY 249 FADSVVSESQTS)LAGSV)SLDES)GHNSM)GSDGT)G)S--DSDENVAARAPPELQ 307
 Db 1109 EKALKEVKEENEHLKEEIKOLEKEATEKQQLNSLANLESLEKEHEDLA-----Q 1160
 QY 308 LRFQMEVAPALQCKNIIICLP---TSGKTRVAVYIAKDHLDKKKKASEYKXVYLVN 364
 Db 1161 LKRYEDQANKERQYNEISQLNDEIT)SQQENESIKKKNDELEGVAKMS----- 1212
 QY 365 KVLVEQLPFKEFQPLKKRYRIG)SGDTOLKISEPEVAKSCDII)STAOI)ENS)NL 424
 Db 1213 -----TSEQSNLKK-----SEIDALNLOIKELKKNE--TNEASILE-----SI 1250
 QY 425 ENGEDAGVQID)SFL)I)IDEC)HTNK)EAY)NIMRY)LMOKLKN-----RLK)NK 476
 Db 1251 K)VESE)TVKIKELQ---DECKNEKEV-----SELEK)K)K)ASEDKSKSY)ELQ)KE 1299
 QY 477 PV-----IP)POLIG)TASPGVGCATK)K)K)EHL)K)LCAN)DA)FI)K)Y)KN 524
 Db 1300 KIEELDAKTEIKIQELKIT)N)SK-----AK)KESELSIK)K)TSSEER--K)N)ED 1350
 QY 525 LDQ)KNOIQ)EPCK)F)ADATREDPFKEKLE)MTRIQ)G)Q)MSP)SDGTQ)PYEOW)Q 584
 Db 1351 LELK)N)EIQ)KNO)AF-----KERKL-----LNBSSITQ)YSEK)NT 1389
 QY 585 MEKKA)K) 643
 Db 1390 L)E)ELR)I)LO)N)E)N)D)-----K)KEID)NT)SE)LE)K)SL)S)N)DE)LE)K)ON)T)K)S)N) 1439
 QY 644 -----DSDEGD)G)E)Y)C-----D)G)DE)D)D)K)K)P)K)D)E)R)P)L)M)T)P)F)EN)K)K)Y) 694
 Db 1440 I)LSYK)D)IT)RNDK)LS)IE)R)K)R)D)LS)E)K)Q)D)LA)Q)ESK)A)Y)-----E)N)IK)LE)ES 1493
 QY 695 -----EY)EN)E)K)IT)K)R)N)T)I)M)D)Y)R)TES)A)G)I)FT)R)OS)A)VA)LS)OM)I)E)N)E)F)A)E)Y 747
 Db 1494 SK)K)A)EL)E)K)E)K)M)K)K)LE)S)IE)N)E)I)EL)K)SS)M)T)I)-----RK)S)K)E)L)E)D)-----SK)S)A)E)E 1544
 QY 748 G)VA)H)H)IG)A)H)S)E)K)P)M)T)Q)N)E)K)E)Y)S)K)F)T)G)K)I)N)L)I)A)T)VA)E)G)D)IK)C)N)I)V)R) 806
 Db 1545 D)K)-----N)LO)HE)K)S)D)LS)R)-----IN)SE)ND)I)E)I)K)S)K)I)P) 1576
 QY 807 -----Y)G)Y)T)E)I)A)M)Y)Q)A)R)G)R)A)D)E)S)T)Y)Y)A)H)S)G)S)Y)I)E)H)E)Y)N)D)F)R)E)M)M)K)A)H) 859
 Db 1577 E)AK)S)E)L)E)T)V)K)O)E)L)N)A)O)K)I)R)I)A)E)N)T)-----V)K)S)K)E)D)E)R)I)K)Q)J)A)E 1625
 QY 860 G)Y)N)M)K)P)E)Y)A)H)-----K)I)E)LO)M)O)S)I)E)K)K)K)T)K 889
 Db 1626 I)K)S)O)E)K)E)L)T)S)R)I)K)E)L)D)E)LD)S)T)O)O)K)O)K)S)E 1658
 RESULT 8
 BRR2 YEAST
 ID BRR2 YEAST STANDARD; PRT; 2163 AA.
 AC P32639;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Pre-mRNA splicing helicase BRR2 (BC 3.6.1.-) (Protein Smu24b).
 GN BRR2 OR RSS1 OR SMU246 OR YER172C OR SYCP-ORR66.

OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes;
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288C / AB972;
 RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
 Aviles E., Berio A., Brennan T., Carpenter J., Chung E., Cherry J.M.,
 Chung E., Duncan M., Guzman E., Hartzell G., Hunkeler-Smith S.,
 Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,
 Mosedale D., Nakahara K., Namath A., Norgren R., Refner P., Oh C.,
 Petel F.X., Roberts D., Sehl P., Schramm S., Shougen T., Smith V.,
 Taylor P., Wei Y., Yelton M., Botstein D., Davis K.W.;
 RA Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
 RL [2]
 RN SEQUENCE OF 1-169 FROM N.A.
 RA Mulligan J.T., Dietrich F.S., Hennessey K.M., Sehl P., Komp C.,
 Wei Y., Taylor P., Nakahara K., Roberts D., Davis K.W.;
 RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP CHARACTERIZATION.
 RC STRAIN=DBY473;
 RC MEDLINE=96304576; PubMed=8722763;
 RA Noble S.M., Guthrie C.;
 RT Identification of novel genes required for yeast pre-mRNA splicing
 RT by means of cold-sensitive mutations.*;
 RL Genetics 143:67-80(1996).
 RN [4]
 RP CHARACTERIZATION.
 RC MEDLINE=96324408; PubMed=8670905;
 RA Lauber J., Fabrizio P., Teigelkamp S., Lane W.S., Hartmann E.,
 Luehmann R.;
 RT The HeLa 200 kDa U5 snRNP-specific protein and its homologue in
 RT Saccharomyces cerevisiae are members of the UEXH-box protein family of
 RT putative RNA helicases.*;
 RL EMBO J. 15:4001-4015(1996).
 CC -1- FUNCTION: REQUIRED FOR PRE-MRNA SPLICING.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potencia.).
 CC -1- DOMAIN: COMPOSED OF TWO SIMILAR DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE HELICASE FAMILY. SK12 SIMILARITY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: U18922; AAB64699.1;
 DR PIR: S30856; S30856.
 DR SGD: S0000974; BRR2.
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR004179; Sec63.
 DR Pfam: PF00270; DEAD_2.
 DR Pfam: PF00271; Helicase_C_1.
 DR SMART: SM00487; Sec63_2.
 DR SMART: SM00487; DEXDC1_1.
 DR SMART: SM00490; HELIC_C_1.
 KM Hydrolase; Helicase; mRNA processing; mRNA splicing; Spliceosome;
 KM Nuclear protein; ATP-binding; Repeat.
 KM DOMAIN 521 923 1.
 FT DOMAIN 1370 1757 11.
 FT NP_BIND 521 528 ATP (POTENTIAL).
 FT NP_BIND 1370 1377 ATP (POTENTIAL).
 FT SITE 634 637 DEH BOX.
 FT SITE 1474 1477 DDAH BOX.
 SQ SEQUENCE 2163 AA; 24618 MW; DFAFEE387168D944 CRC64;
 Query Match 3.9%; Score 204.5; DA 1; Length 2163;
 Best Local Similarity 19.1%; Pred. No. 0.011;

Matches 162; Conservative 136; Mismatches 247; Indels 40; Gaps 40;
 QY 127 DKLVAVDVKME-----ELTIED-----RNR--IAAR 156
 DB 1305 DTSVDDISEKILINDIETLEHNPVLEQVLQVLLKFNLSIAEFLTKNRSTFWKILPLAK 364
 QY 157 NNGNEGVALLKRIYOKNNMFSAFLNVLROGNNELVOLGSGCSSESNATLENSVD 216
 DB 365 STENE--IPNLLEKVVAK-----GLNDLVEQYKFRRTTSKRITD--SGDD 406
 QY 217 GROVE-----OLLSTVQNLKEVWGMENNESSPADS--VSESPTSLAVSGSTED 271
 DB 407 QGSSSAKRTKSNRAIPVIDLE-----KIKFDESKLMTVTKSLVDSGPK-- 454
 QY 272 ESLGHSNNGSDSGTMSDSDENVAARASPEPELDR----- 309
 DB 455 -----RVKPYDEIHIPASKPIVDLEKETSLLPMQCEAFSTSTSLN 500
 QY 310 PYQMEVAQPALEG-KNIITCLPTGSGKTRVAVY-----IAKHLDKKKASIPKVIYV 363
 DB 501 PLSQKVFIAAFEGDSNMLICAPTGSGKTNIALTVLKALSHYNNPKTKLNISAKIYVI 560
 QY 364 N--KVLVLEQLPRKEFO--PELKKWRYVIGSGDTOLKISPEVVKSCDILITVAJLE 418
 DB 561 APKLALVQGV--REGRRLAFL--GIKVALTGDSRLS--RKQIDETQVVSITPEKMD 613
 QY 419 NSLLNLNENGDAVOLDSFSLIIDECH--HTNKEAVYNNIM-RHYLMOKIKNKILKEN 475
 DB 614 ITRNSNN--LAIVEL--VRLIIDEHLHDGPGVLESIYARFMAKSKYQV-- 663
 QY 476 KVIYPLPOLGISTASGVGATKQAKAEHILKCANIDAFITIKYKENVLQKMLIEP 535
 DB 664 -----YPRILIGLSAT-----LPNVEDGRF-LKVPKGLFLFDSNFR-P 700
 QY 536 CKKFAIDATRDPEKELLETIRIQTGYCQSPSPSDCTQPEQWALOMEKKAKKGNR 595
 DB 701 C-----PLS-----QPPGCIKENSIKNIKAKA 721
 QY 596 KERVQAEH-LRRYNALQINDTIRMIDAVTILETYNEBKQKFAVIDEDSDGMDHYC 654
 DB 722 NMDACEYKVLSEINQNL--IVFVSRKET----- 750
 QY 655 DODEDEDLKKLKLDEIDRFIMTLTFENNNKKLKLAEPEYENKLTKLKNTIMHYTR 714
 DB 751 -----SRATWLNKFAEN-----ITIKLT 772
 QY 715 TESASGLFTKROSAYALSQWITENEKFAVGVKAHHLIGAGHSSEFKPTQWJUREV 774
 DB 773 NAGSKQILKT--EAANYLDPSL--KRLIESGIGTH--AG-----ITRSWISL 816
 QY 775 ISKFRGKINLLIATVAEGLDIKECNVIRYGLV-----TNFIAWYQARG 821
 DB 817 EDLFAJGLIQLVATATLAWGVNLPANHVIYLIKQDVYSPKSGWBSLPQVLDQWLAG 876
 QY 822 RARADESTYVLAHSGSGVIEHETVNDREKKMYAALHCYQNMKREYFAHKLIEIUMQSI 881
 DB 877 RPRYD--TF-----GEGII-----ITDSNVQYVLSVNOQPIESQPVSKIVINIAEV 924
 QY 882 MEKKMKTKRN 891
 DB 925 VAGNIKCRND 934
 RESULT 9
 SBCC_CLOAB
 ID SBCC_CLOAB STANDARD; PRT: 1163 AA.
 AC Q97FK1;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Nuclease sbccD subunit C.
 GN SBCC OR CAC2736.
 OS Clostridium acetobutylicum.

CC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 CC Clostridium.
 CC NCBI_TaxID=1488;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
 RA MEDLINE-21359325; PubMed-11465286;
 RA Noelling J., Breton G., Omelchenko M.V., Makarov K.S., Zeng Q.,
 RA Gibson R., Lee H.M., Dubois J., Qiu D., Hittl J., Wolf Y.T.,
 RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
 RA Bennett G.N., Koonin E.V., Smith D.R.;
 RT Genome sequence and comparative analysis of the solvent-producing
 RT bacterium Clostridium acetobutylicum*;
 RL J. Bacteriol. 183:4823-4838(2001).
 CC - FUNCTION: sbcd cleaves DNA hairpin structures. These structures
 CC can inhibit DNA replication and are intermediates in certain DNA
 CC recombination reactions. The complex acts as a 3'-5' double
 CC strand exonuclease that can open hairpins. It also has a 5'
 CC single-strand endonuclease activity (By similarity).
 CC - SUBUNIT: Heterodimer of sbcd and sbcd (By similarity).
 CC - SIMILARITY: BELONGS TO THE SMC FAMILY. SMC SUPERFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AEO07771; AAK0682.1;
 DR InterPro; IPR00439; ABC_transport.
 KM Hydrolyase; Nuclease; Exonuclease; Endonuclease; DNA replication;
 KM DNA recombination; ATP-binding; Coiled coil; Complete proteome.
 FT NP_BIND 35 42 ATP (POTENTIAL).
 FT DOMAIN 197 415 COILED COIL (POTENTIAL).
 FT DOMAIN 446 1003 COILED COIL (POTENTIAL).
 FT SEQUENCE 1163 AA; 135507 MW; CEF50BD21507A92 CRC64;
 Query Match 3.7%; Score 197.5; DB 1; Length 1163;
 Best Local Similarity 18.8%; Pred. No. 0.011;
 Matches 197; Conservative 187; Mismatches 999; Indels 263; Gaps 45;
 QY 31 VLDVLTLPFAVEKQIQRTVATSGNMQAVELLSTLEKGVHIGTREFVEALRTGSP 89
 DB 44 ILDSITLSLGEVARKSNFNTNCSLNSVFEQISGKELKLVREFF---RKDNKT 99
 QY 90 LAARYNPELTDLPSPFENAHDEYQLNLLOPTLVKLLVSDVDLCKMPELITTEDR 149
 DB 100 GSVRSKSAKIVDITGDEVEYLEG-----AKSVNKKCCTIGLSLDDF 142
 QY 150 NRMAAEN-----NGNSSVRELLKRIYQKEM---PSALINVIHQ-TGNNELV 194
 DB 143 TRIVVLPQGRSEFLKLEGRER--RNMLERFLNDEYGDLSFLAKRIKEREKENLV 200
 QY 195 QELTG-----SDCSNAEIEINLSQVDPQVEQLLSTTVQPNLEKEMVNNSSSP 249
 DB 201 GELGYNENINDVAKERRELLKNNDFNENSKYELAAEEYNKREWMVQIETIEKNR 260
 QY 250 ADSSVSESSTSLAEGSVCLDESGLHNSNMGSDSGTNGDFFNVAFAASPEPEQLR 309
 DB 261 VRKDLMEKKD-----EIDLRKRAHLIG-ESSSKVK 289
 QY 310 PY-----QMEVAPALEGKNIITCLPTGSKTRVAVYIADNLDKRRKASEPQVY 360
 DB 290 PYIDYENTLKQIDILKEQLISRE-----NTMKALISLEKFMMEKRI STAKDKRK 339
 QY 361 VL-----VAKVLLVEQLRFRKEFQPLK-KMYRVIGLSGDTOLKIFPPEV---VKSDDI 410
 DB 340 ALPKFMKIHHTIIDAIEKELDLNKKLEKRLQKIEKISLEASNKELIKONIKDIDL 399
 QY 411 ISTAQLIENSLNLNENGEDAGVOLSD--FSLIIDE-CHHTN----- 449

DB 400 TLKIQMLESKIDNLKVPPEEYKKNINIEGIFLBNRYDEKLHKKNKGLGDCDFVETKAS 459
 QY 450 -KEAVNNIMRHLYLMOKLKNRNLKKNKRVYLPOLIGLTASGVGA-TKQAK-PHHL 507
 DB 460 KKEMLFNKLEERSKIDYTKRLQDLNKKD-FPKDVL-LTFOKLDSQKQAKISENE 517
 QY 508 KLCANDAE-----TIKTVENIDOLKNQIO-----EPCKKPAIDAEARE----- 547
 DB 518 SLKASLRVENSQVLRKTKEKTKLEDKISKVNKINIESLETENMAHYLRKIKSEACP 577
 QY 548 -----DPFEKLEIMTRIQTQVQSPMSDF-GTQPYE-----QMA-TTKKAA 590
 DB 578 VCGSVHIRKGFREVDLKALETLK-----SELGFKKRFEENETVMEASIKVEENK 632
 QY 591 KKN-----KREVCACHELRKKN-ELAQINDIR-MTAVHILETF 629
 DB 633 KKLNESINNLGEEFKVSLSEMEKKFNVLKRYKKNLKLQDLNKTOLSEKSKIEVE 692
 QY 630 YNEEK-----DKRFVIEDSDGDDDEYDCGDEDEDLKLPLKLDETDFLM;LFFEN 683
 DB 693 YQKEKIVEKQEKRIYDLKSELEE-AIKFENEVATIENTKALKLQD-----PKFEM 744
 QY 684 NKMLKRLAENPEYENKELTKRLNTIMEQYTRTESARGIIFTKROSAYALSOW;IENK 743
 DB 745 KEILKEKRYVFAEGE-IKDLRLNLNIRHTEKQ-----LMDKCSRLKEI-KKN- 793
 QY 744 FAVGVKAHHLIAGHSSEKPRPTONQOKVIAKFRGKINLLIATVAEFGIL-KKCN 803
 DB 794 -AEL-----KKDKIINEKIELI-KNVGVGLDNLVELKEKIEG-TIKK-- 834
 QY 804 VIRGLVTNEIANVQARGRADESTYVLAHSGSVIEHEETVNDREKMYKAHVCN 863
 DB 835 EEOYNLDCKKMNLEDKYKCSDE---LIKYSNLSSLDKRNKD-----IKLKK 882
 QY 864 MKPEEYAHKLELDQMSIMEKMKKTKRNIAKHYKNNSPLTFLCKKCSVLACS; 1HWI 923
 DB 883 ILMEEFENIEKAKENVILNKLKSDVEKYNELSKV-----NGAVEVLS- 930
 QY 924 EKMHHVMPEFKELIVRNKALQK 949
 DB 931 KKLKRNKL--EKKWLEIQQNNRVER 953
 RESULT 10
 G064_HUMAN STANDARD; PRT; 2230 AA.
 AC Q13439; Q14436; Q13270; Q13654;
 DT 16-OCT-2001 (Ref. 40, created)
 DT 16-OCT-2001 (Ref. 40, last sequence update)
 DT 15-JUN-2002 (Ref. 41, last annotation update)
 DE GOLGI autoantigen, golgin subfamily A 4 (Trans-golgi P230) (56 kDa
 DE golgin) (Golgin-245) (72.1 protein).
 GN GOLGA4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RA MEDLINE-96215236; PubMed-8626529;
 RA Eilich R., Gleeson P.A., Campbell P., Dietzsch E., Toh B. H.
 RT Molecular characterization of trans-golgi P230: a human periphera
 RT membrane protein encoded by a gene on chromosome 6p12-22 contains
 RT extensive coiled-coil alpha-helical domains and a granin motif.
 RT J. Biol. Chem. 271:8328-8337(1996).
 RL [2]
 RP SEQUENCE FROM N.A.
 RA Seelig H.P.;
 RL Submitted (NOV-1994) to the EMBL/Genbank/DBJ databases.
 [3]
 RP SEQUENCE OF 131-2230 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE-96125112; PubMed-8537393;


```

QY 557 -IMRIOTYCOMSPMDGTOFY-----EOMLIONEK-----KAAKKNH-KRV 599
Db 305 GFGSGIENYSHLIRPLGSLPTLLDFFGDMVLMIIDESHVTL:POYRGMTNNS-AKV 364
QY 600 CAEHLRKNEALQINDTIRMLDAVTHLETFYENNEKKDKFAVLEDDSEGDHYVTEETED 659
Db 365 LVHDGFRPLSALD-NRPLK-----PEEF--EKKTQLVYV--SAPFG-- 401
QY 660 EDDLKPLKLDDETFMLMTLFFENKNMKLRLAENPEVNEKLTILRLNTIMCYIHLISA 719
Db 402 -----PEEIHETDKMWOIIRPTGLDLPKLEVRPT-ENO-----IDDLSEIQLVINE 450
QY 720 RGLIFTKROSAYALISOWITENEFKFAEYGAHHLIGAGHSEPKRPMTONOKVIFER 779
Db 451 RVLVTLTLTKMSSEDLITM-----KEAGIVNVL-----HSEIKTL--ERIEIHLRL 496
QY 780 TGIINLLIATTVAEGLDIKECNVI-----RYGLVNIEMAVOAGR-ARADESYVLV 833
Db 497 MGTVDVIVIGMLREGIDIPESVSLIIDLADKEGLFHSRSLIOTIGAAKNDKRVNRY 556
QY 834 AHSOGS-----VIEHETVNDREKMYAIH-----CYONKPEVYAHKI 873
Db 557 ADKMTDSKKVYIDETQRREIOMKNEKHGILPTKINKIHDLISATVENJFNHKAQTV 616
QY 874 LELQMOSIMEKMKTKRNIAHXY 897
Db 617 IPKKMTK--KEROKTIDNIEKMK 638

RESULT 12
MYSN_DROME
ID MYSN_DROME STANDARD: PRT: 2017 AA.
AC 099323:
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, non-muscle (2ipper protein) (Myosin II).
GN ZIP.
OS Drosophila melanogaster (Fruit fly).
OC Euarthropota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
RN NCBI_taxid:7227;
RP [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=90349606; PubMed=2117279;
RX Ketchum A.S., Stewart C.T., Stewart M., Kiehart D.P.;
RT "Complete sequence of the Drosophila nonmuscle myosin heavy chain
RT transcript: conserved sequences in the myosin tail and differential
RT splicing in the 5' untranslated sequence."
RL Proc. Natl. Acad. Sci. U.S.A. 87:6316-6320(1990).
CC -1- FUNCTION: NONMUSCLE MYOSIN APPEARS TO BE RESPONSIBLE FOR
CC CELLULARIZATION. REQUIRED FOR MORPHOGENESIS AND CYTOKINETIC.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN
CC -1- SIMILARITY: CONTAINS 1 TO DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (see http://www.isb-s.b.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M35012; AAA28713.1; -
CC PIR: A36014; A36014.
CC PIR: B36014; B36014.
CC HSSP: P10587; 1BR2.
CC Flybase: FBgn0005634; zip.

```


DR InterPro: IPR000048: IQ-region.
 DR InterPro: IPR004008: Myosin_N.
 DR InterPro: IPR002928: Myosin_tail.
 DR InterPro: IPR001609: myosin_head.
 DR Pfam: PF00063: myosin_head. 1.
 DR Pfam: PF00612: IQ. 1.
 DR Pfam: PF01576: Myosin_tail. 1.
 DR Pfam: PF02736: Myosin_N. 1.
 DR PRINTS: PRO0193: MYOSINHEAVY.
 DR PRODOM: PD000355: myosin_head. 1.
 DR SMART: SM00015: IQ. 1.
 DR SMART: SM00242: MYSC. 1.
 DR PROSITE: PS50096: IQ. 1.
 KM MYOSIN: Alternative splicing: Coiled coil: Actin-binding:
 KM ATP-binding: Calmodulin-binding.
 FT DOMAIN 1 829 MYOSIN HEAD-LIKE.
 FT DOMAIN 830 859 IQ.
 FT DOMAIN 886 2017 COILED COIL (POTENTIAL).
 FT NF_BIND 225 232 ATP.
 FT DOMAIN 250 260 25 KDA/50 KDA JUNCTION.
 FT DOMAIN 682 694 50 KDA/20 KDA JUNCTION.
 FT DOMAIN 705 727 ACTIN-BINDING.
 FT DOMAIN 742 758 REACTIVE SULFHYDRYL/ACTIN-BINDING.
 FT DOMAIN 1303 2017 LIGHT MECHANOSENSING (LMM).
 FT DOMAIN 1303 1970 ALPHA-HELICAL TAILPIECE (LMM).
 FT DOMAIN 1971 2017 GLOBULAR TAILPIECE.
 FT VARSPLIC 1 45 MISSING (IN SHORT ISOFORM).
 SQ SEQUENCE 2017 AA: 232016 MW: 738302.84 kDa: 2528 GRC64:

Query Match 3.6%; Score 193.5; DH 1; Length 2017;
 Best local similarity 18.3%; Pred. No. 0.037;
 Matches 206; Conservative 182; Mismatches 186; Indels 351; Gaps 45;

DB 83 L-----RRTGSPLAARYNNPELTLP-----SSEFNHDFYLT--- 117
 DB 930 LVKRTLADELOAEIELCAEAEESRRLMAR--KOELEIMOELETRIEERIVATLGG 987
 QY 118 -----LNL-----LOPTLVKLLVLDVLD-----KMEPELITEDS NRIAAFNMM 159
 DB 988 EKKLELNIQDLEDEGLEEEEAHQKQLEKVOADAKIKKEEDLALTDQ NQKLLKERRK 1047
 QY 160 NESGVBELKRIYOKENMF--AFNLVLCOTGNELVQEL-----TQSGSESNAVL 209
 DB 1048 LEEBRANDLSQTLAEEBEKAKHLAKKAKHEATITTELEERLHKDQOCORPSMSKRIETE 1107
 QY 210 -----ENLSQVDPQPV-----EQLLST----- 227
 DB 1108 VADLKQLNERRVQVDEMQAQLAKKEBELQTLIRIDHSATKATQKKA RELESQIAET 1167
 QY 228 -----TVQPLEKEVGMEN-----NSSESSPAUSSVSESLSAEQSVSG 269
 DB 1168 QEDLEAKKARAKAEKVRRLDSELEALKNELLSLDTTAAQELSKSKFCCLATILKSL 1227
 QY 270 LDESLOHNSMG-----SDSGTMSQSD--ENVAARASPEPELQRLPYQFVAQPALEGN 324
 DB 1228 EEEETVHNEGVLDMRRKHSELEMSINDOLEN-----LRKATVLEKKGTLIAEN 1277
 QY 325 IILCLPTGSGKTRVAVYIAKDHDKKKASPEQKIVLVNKKVLLVQQLFKKKEFOPLAKW 384
 DB 1278 ADLATELRS-----VNSSQENDRRRQAE--SQIAELQVKAELIERA--KSEIQ----- 1323
 QY 385 YRVIGLSGDTOLKISFEVAKSCDIIISTAQIENSLNLENDIAGV-----QLSD 436
 DB 1324 -----EKCTKQGAENITNOLBEAELEKASAAVVSASMSQSLTE 1363
 QY 437 FSLIITIDEGH-----HTNKEAVYNNIMRRHYLMOKLNNELKKEFKPVLPLOI 484
 DB 1364 AQQLEEFTRQKLGISSKLQIJESEKALQEOLEED--DEAKRNERKRLAVTQOMQEI 1420

QY 485 LGITLSPGVGATKQAKAEHILKLCANLDAFTIKTVKENLDOLKNQIO--EY NKKFALA 542
 DB 1421 -----KKKAE-----DADLAKLEEGKKRLNMDLEALIK VKHIA 1457
 QY 543 DATREDPEKKEL--LEITRIOTYCOMSPMSDFGQPEQIAQIOMEKKAASKNKKI RV 599
 DB 1458 QNDRLQSKKKKIQSELEDAI--LELEAQRTVLELEKKQKNPKITAEKALISQIAQIRD 1516
 QY 600 CAENLRKYNALQINDTIRIMDAYTHLETIFYNEBKDKKFAVIEDSDDEGDEHY IDCHD 659
 DB 1517 TEREAREKETVYLSRSRLDEAFDKIEDLEKKR-----TQNELDOLANTQ--GTAD 1566
 QY 660 EDDKLKPLKLDLDRFLMLTFEENKMKLRLAENPEYENE-KLT--KIR-----NTIME 710
 DB 1569 ---KNVHELEKAKRAL-----ESQLELKQNELEDEDLQITEDAKLRLVNNQALRS 1618
 QY 711 QYTR-----TESARGII-----FTKTROSAVALSOWITENKIA AFGVK 750
 DB 1619 OFEROLLAKEEBAEERKQVLVQRLDLETLEDEKQRTAAVASKKKEGDKI ETTME 1678
 QY 751 ANHLI---GAGHSSEKPTON-----DQKEYISKFRGKINLILATITVAEG 795
 DB 1679 MNKKVQEDALKKAKKLOAQVQKALDQAEAKAKAEIQLALSEADQKVALFAVLO--- 1735
 QY 796 LDKECNIVIRIGLVTNEIAMVQAGRAADSTVYLVASHSGVIEHETVNDREKMMY 855
 DB 1736 -----LTEDLASSE--RARRAETEREDELAEELIANNANKSLMDIKRLEA 1780
 QY 856 KAIHCQVQNKPREVYAKIL-----ELQMSIMEKKKTKNKIAKKNNNSI FIFLCK 908
 DB 1781 RALTELEELFEQDSSEVLLDRAAKAROLQIETTELANKSNSQKN-ENGANILR --- 1835
 QY 909 MCSVLACSGEDIHVIEKMHVNMTPFEKELYIVRENKALQKKCADYQ 955
 DB 1836 -----RONKELKALAEIE 1844

RESULT 13
 HELLS_PYRAB
 ID HELLS_PYRAB STANDARD: PRT: 715 AA.
 AC Q9Y0A9;
 DT 16-OCT-2001 (Rel. 40, created)
 DT 16-OCT-2001 (Rel. 40, last sequence update)
 DT 16-OCT-2001 (Rel. 40, last annotation update)
 DE Putative sk12-type helicase (EC 3.6.1.1).
 GN PAB0592.
 OS Pyrococcus abyssi.
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
 OC Pyrococcus.
 OX NCBI_TaxID-29292;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-GE5 / Orsay;
 RA Hellig R.;
 RT "Pyrococcus abyssi genome sequence: Insights into archaeal chromosome structure and evolution."
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 CC -1 SIMILARITY: BELONGS TO THE HELICASE FAMILY. SK12 SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.1st.sib.ch/annotation/> or send an email to license@sib.ch).
 CC
 CC EMBL: AJ248285; CAB49795.1;
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR003583; HHH.1.
 DR InterPro: IPR001650; Helicase-C.
 DR InterPro: IPR000445; HHh.
 DR Pfam: PF00270; DEAD; 1.

[illegible]

Db 2248 OKKSTTRLOELBQENKLFKDMKELGL-----AIKESQAMSTODQVVL---FGKRA 2296
 QY 516 FTTKTKENLDOLKNOIOECKKFAIADATREDPFKEKLL-ELMTIOTVYQMSPSMDRG 574
 Db 2297 QIIIOEKREVIDQNEQVTKIOOOLKITTDKVIIEKNEELHLEIOTIE-CLMS----- 2248
 QY 575 TQYEDMAIQEKKAAKKGRKRCVCAEHKRNKNEALQINDTRMDATVHLETFYNEEK 634
 Db 2349 -----DOECVNR-NREE-----ETPQINEVYI-----EKL 2371
 QY 635 DKRFAYIEDSDGDEYCDGDEDEDLKPKPLKLDLTDLTFEENKMKRLAENP 694
 Db 2372 QOELANIGOKTSMNANSL-----SEPADSLKHQDVIYAEKL-----ALIEQV 2414
 QY 695 EYENKLLKLRNTIMEQYRTESANGIIFTKROSAYALISOMITENEKAEVGVKAHLL 754
 Db 2415 ETANEEMTEKKNVLEKTFENFQNLTOELFSLKRESEVEKIQSIPEN---SVNAVADHL 2470
 QY 755 IGAGHSEEPK-----MTONEKEVISK--FRT-----GK---IN-----LITVAEEG 795
 Db 2471 -----SKORPELEVLTEDALKSLKQVYFKSPENGKSTINLETRLQLESTVSAKD 2524
 QY 796 LDIKEC-----NIVIRGLVTEIEM-----816
 Db 2525 LELTQCKQIKDMQEQFETEMLOKKIVNLOKIVERKVAALVSOIQI-ANQEVAKFCQ 2584
 QY 817 -----VOARGARADE-----STYLVVAHSGGVTEHETVND 848
 Db 2585 DNOITSEPERTIOMLNQREBELGSDISALTLRILESESQVEMHT-SLEIEKEQV-E 2642
 QY 849 FREKMYKAIHCYQNNKPEYAKHILELOMSIMEKMKTKRNIAKYNKNPDLITFLCK 908
 Db 2643 IAEKNVL-----EKKKLEL-OKLEGENEKQRE--EKKKSP-----2678
 QY 909 NCSVLACSGEDIVHVI-----EKMHVNTTPFKELIYVR-ENKALOKKCAVQINGELICK 963
 Db 2679 -----QDVEYKLTTELPHSNESGFFNELEALRAVSATKAEIAYEKAE-----2725
 QY 964 CGQAMGTMYHKGLDPLCLIRNFVYFKNNSTKKQ 999
 Db 2726 -----KLOELLIVKETNMTSLQ 2742

RESULT 15
 CENT_HUMAN STANDARD: PRT; 3210 AA.
 ID CENT_HUMAN STANDARD: PRT; 3210 AA.
 AC P49454; Q13246; Q13171;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE CNP-F, kinetochore protein (Centromere protein F) (Mito-sin) (AH antigen).
 GN CNP-F.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eumetazoa; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; H. mo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Breast carcinoma;
 RX MEDLINE=95348175; PubMed=7542657;
 RA Liao H., Winkfein R.J., Mack G., Rattner J.B., Yen T.J.;
 RT "CNP-F is a protein of the nuclear matrix that assembles onto kinetochores at late G2 and is rapidly degraded after mitosis.";
 RL J. Cell Biol. 130:507-518(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95379848; PubMed=7651420;
 RA Zhu X., Mancini M.A., Chang K.-H., Liu C.-Y., Chen C.-H., Shan B., Jones D., Yang-Feng T.L., Lee W.-H.;
 RT "Characterization of a novel 350-kilodalton nuclear phosphoprotein that is specifically involved in mitotic phase progression.";

RL Mol. Cell. Biol. 15:5017-5029(1995).
 RN [3]
 RP SEQUENCE OF 2194-3210 FROM N.A.
 RX MEDLINE=95336446; PubMed=7612011;
 RA Li Q., Ke Y., Kapp J.A., Fertig N., Medsker T.A. Jr., Joshi H.C.;
 RT "A novel cell cycle-dependent 350-kDa nuclear protein: C-terminal domain sufficient for nuclear localization";
 RL Biochem. Biophys. Res. Commun. 212:220-228(1995).
 RN [4]
 RP CHARACTERIZATION.
 RX MEDLINE=95370296; PubMed=7642639;
 RA Zhu X., Chang K.-H., He D., Mancini M.A., Brinkley W.R., Lee W.-H.;
 RT "The C terminus of mitosis is essential for its nuclear localization, centromere/kinetochore targeting, and dimerization.";
 RL J. Biol. Chem. 270:19545-19550(1995).
 RN [5]
 RP CHARACTERIZATION.
 RX MEDLINE=98437347; PubMed=9763420;
 RA Chan G.K.T., Schaar B.T., Yen T.J.;
 RT "Characterization of the kinetochore binding domain of CNP-E reveals interactions with the kinetochore proteins CNP-F and hBUBR1.";
 RL J. Cell Biol. 143:49-63(1998).
 CC - FUNCTION: PROBABLY REQUIRED FOR KINETOCHORE FUNCTION, INVOLVED IN CHROMOSOME SEGREGATION DURING MITOSIS. INTERACTS WITH RETINOBLASTOMA PROTEIN (RB), CNP-E AND BUBR1.
 CC - SUBUNIT: HOMO- OR HETERODIMER.
 CC - SUBCELLULAR LOCATION: NUCLEAR MATRIX (BUT NOT IN THE NUCLEOLUS), REORGANIZATION TO THE KINETOCHORE/CENTROMERE (CORONAL SURFACE OF THE OUTER PLATE) AND THE SPINDLE DURING MITOSIS.
 CC - DEVELOPMENTAL STAGE: GRADUALLY ACCUMULATES DURING THE CELL CYCLE.
 CC - PTM: HYPERPHOSPHORYLATED DURING MITOSIS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 DR EMBL: U19769; AAA82889.1; -
 DR EMBL: U30872; AAA82935.1; -
 DR EMBL: U25725; AAA68889.1; -
 DR Genew: HGNC:1857; CNP-F.
 DR MIM: 600236; -
 DR InterPro: IPR001230; Prenyl-site.
 KW Chromosomal protein; Centromere; Coiled coil;
 KW Mitosis; Phosphorylation; Antigen; Cell cycle; Repeat.
 FT DOMAIN 14 197
 FT DOMAIN 273 769
 FT DOMAIN 823 1328
 FT DOMAIN 1642 1746
 FT DOMAIN 1862 2987
 FT DOMAIN 2207 2568
 FT REPEAT 2207 2386
 FT REPEAT 2389 2568
 FT DOMAIN 3015 3032
 FT CONFLICT 16 16
 FT CONFLICT 250 250
 FT CONFLICT 272 272
 FT CONFLICT 611 611
 FT CONFLICT 1494 1589
 FT CONFLICT 1611 1611
 FT CONFLICT 1811 1811
 FT CONFLICT 2242 2243
 FT CONFLICT 2335 2335
 FT CONFLICT 2492 2492
 FT CONFLICT 2545 2561
 SQ SEQUENCE 3210 AA; 367589 MW; 11DB3324960E4334 CRC64;
 Query Match 3.6%; Score 189.5; DB 1; Length 3210;
 Best Local Similarity 19.1%; Pred. No. 0.11;

Matches 218: Conservative 187: Mismatches 473: Indels 265: Gaps 46:

```

OY 21 RAKMYIQVEBVL-----DYLTFLPAEYKEQIQFVATSG-----NMQAVELLIS 64
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1917 KIEACIELEKIEGELAKKENSIDSEKLEYFSCDHQELQIVETSEGINSDI DMHAKSSKE 1976
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 65 TLEKGYMHL--GWTREFVELARTGSPLAARYNNPELTJLPSRSTFNAMH IYLOJLNLIQ 122
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1977 DIGDNVAKYNDKSERFLD-----VENLSRIKSEKASIEH-IALYLFADLE 2022
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 123 PTLVDEL-LVRDYLDK-----GMEELLTTI-EDRNRIIAAENNGNE--S-VARELLKRIYQ 173
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2023 VVGTEKICIEKEMENKQXIVYCLEEELSVYTSERNQKRGELDTMSKKTIVILQJLSKMMK 2082
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 174 K----ENMFSAPLNLVLRQT-----GNNELVQELTSGDCNSNAELENLSQ VDGPOVEEQ 223
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2083 KIOELESHOSSECHICIQVAAEAYEKEKTELLQTLIS-SDVSELKDKTHLQF KIOSLEKOSQ 2141
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 224 LSTT-----VOPNLEKRYWGMENNS-----SESSFADSSVYSRSDISLAEGS----- 266
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2142 ALSLTGCLELNOJAOULNKKERELVKESESLQARLSSESDYFKLNVSKALFALVEKGEFAL 2201
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 267 -VSCLEDSLGHSNMGSDSGTMSDDE-----ENVARASPEPELQI RPYOMEVAVP 318
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2202 RLSSSTQGEV-HQRLRGIEKLVRILEADEKKQHLIAEKLNRKKNDSLKI KVENLFRLEQ 2260
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 319 ALEGNIIICLPTGSGKTAVAYIAK-DHLDKKKASEP-KYIVLVNKKVILVEQLFKKEP 377
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2261 MSEEENQELVLDENSKAEVETLKTQIEEMANSKLKVEELALVLRSEKENI FKIOIEKKG 2320
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 378 QPFLKMYNVIK-----LSGDTQOLKISFPEVYKSCDI IISTAVJLENSLNLNEN----- 426
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2321 Q--LSELDKLLSFKSLLEKEBOAEIQIKEESK-----IAYEMQNOQLPLNEAVAYALG 2372
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 427 GEDAGVQLSDFSLLIIDECHTKEFAVYNNIMHLYLMOKLKNRKLKKNKIVILPUPQ-- 484
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2373 GQGEIIMKATBQSLDPRIEEHQLRNSI-----EKLRARLEADEKKQCLVLOQLKE 2422
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 485 -----LGLTASPGVGATKQAKAREHTILKT-----CANLD 514
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2423 SEHHADLLKGRVENLERELEIATNOEHALLAEENSKGHEVETLAKIEGVOQLRGLEHD 2482
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 515 APTIKTVKNNLDQKNOIQEPCKKFAIADATREDPPREK LELNTRIQTYQOMSPMSING 574
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2483 VWTIRSEKEDLTNELOKEQERISELBI INNSPENLQEK -----EQRKYMKKESSTA 2535
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 575 TOPPEQMAIQEMEKKAKNRKEKRYCA--EHLKRYNEALJIN--DTIRMLDA-----YTHL 626
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2536 MEMILOTOLKELNRYVALINDQFACAKAQONLSOVEGLTLEKAOLOLOI IBAKKNYIVL 2595
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 627 ETTPYN-----EKKDKKFAVIEDDSDEGDDH/GCDDQDEDEDI KKPFLKLETD 673
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2596 OSSVNGLIQVEDGKOLKEKDEISRLKNQI-----QDOEOI VSKLSQVGE 2643
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 674 RLMLTLFFENNNKHLKRLAENPEY-----ENKLTILKLTIMEOYTRTPHSARGI IFTK 726
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2644 H---QIMKQONLELRLVLELQKIQVLOSKNASLQDTLEVLOSSTYKNLENE--LELTRK 2697
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 727 TROSAAVALSQMITEKKEFAVGVKAHHLIGAGHSSEFKPMTONEQKEVLSKFRGTG INLL 786
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2698 MDKMSFV-----EKVNMKTAKETELOREHMEHMAOKTAELOLDELSGKKNRLAGHLOLL 2749
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 787 -----IATVAEGDLDIKECNIVIRYGLVTNEI AMVOAGKARADESTIV 831
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2750 LLEIKSKQOLKELLLENSELKSLDQMHKDOYEKKGKREEIAEYOULRIHEAEKKHQAL 2809
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 832 LVAHSSGVILEHTVNDPFEKMMYKAIHCYQNNKRP-----FYAKHILELOMSIME 883
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2810 LLDTNQYVEVLOT--YREKLITSKE-ECLSSQKLEIDILKSSKEELNNSIKATTO-IIE 2864
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 884 KKKMKT-----RNIAKHYKNNPSLITFLCKNSVLAQSGEIIHYIEK----- 925
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2865 ELKKTMDMLKLYNOLKKEKNERAOGKMLILNSKOLE---EKKELQKHLISOLQAAQEK 2921
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

OY 926 -----MHVNMTPERKELYIVREKKA-----LQKKCADYQINGEITIKKVAAGCTM 971
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2922 QKTGTVMQTKVDELTTETIKELKELTELEKTKADEVLDKYS-LIISHKELPKAK MLIQ 2980
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 972 MVH 974
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2981 VAH 2983
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: May 8, 2003, 15:53:32
 Job time : 51 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 15:50:13 : Search time 52 Seconds
(without alignments)
1894.957 Million cell updates/sec

Title: US-09-515-363c-2

Perfect score: 5311

Sequence: 1 MSGXSTDENFRLSCFRA.....LPITPNLDSECTLESDRD 1025

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR73:*
2: PIR1:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	687	12.9	1037	2 T32534	hypothetical prote
2	654	12.3	811	2 T30968	helicase homolog C
3	642.5	12.1	956	2 A88708	protein COL10.1
4	516	9.7	398	2 T46312	hypothetical prote
5	499	9.4	1134	2 T20332	hypothetical prote
6	453	8.5	752	2 D75219	ATP-dependent RNA
7	441.5	8.3	650	2 D71203	probable ATP-depen
8	417	7.9	741	1 A69432	ATP-dependent RNA
9	402.5	7.6	1374	2 S62524	probable RNA helic
10	392.5	7.4	784	2 H64487	5'-4A family prob
11	350	6.6	1822	2 S44849	K2H4.6 protein -
12	342	6.4	738	1 E69055	ATP-dependent RNA
13	313.5	5.9	993	2 D84386	probable RNA helic
14	303.5	5.7	784	2 S48436	ATP-dependent RNA
15	284	5.3	502	2 T31323	probable RNA helic
16	278.5	5.2	839	2 T39190	ATP-dependent RNA
17	223	4.2	557	2 G69958	probable ATP-depen
18	218	4.1	1979	2 T31622	SNF2 helicase homo
19	217.5	4.1	2663	1 S28261	hypothetical prote
20	214.5	4.0	1790	1 S67593	transcript protein
21	209.5	3.9	449	2 C71156	probable helicase
22	205	3.9	453	2 B75006	RNA repair protein
23	205	3.9	967	2 D72308	conserved hypotet
24	204.5	3.9	2163	2 S50675	pre-mRNA splicing
25	204	3.8	1939	2 T18372	repeat organellar
26	201.5	3.8	1063	2 T38420	probable DNA helic
27	198.5	3.7	2253	2 T30336	nuclear/mitotic ap
28	197.5	3.7	1163	2 G97236	ATPase involved in
29	195	3.7	1531	2 T48946	hypothetical prote

30	195	3.7	3259	1 A56539	giantin - human
31	194.5	3.7	986	2 E90596	res-tion-modifi
32	194.5	3.7	1256	2 T26101	hypothetical prote
33	194	3.7	1039	2 E72734	hypothetical prote
34	193.5	3.6	663	2 G89848	extinctase ABC su
35	193.5	3.6	2017	1 A36014	myosin heavy chain
36	193.5	3.6	2057	2 S61477	myosin II heavy ch
37	193.5	3.6	3225	2 I52300	giantin - human
38	193	3.6	715	2 B75135	DNA helicase relat
39	193	3.6	986	2 H90565	res-tion-modifi
40	192.5	3.6	718	2 E71100	hypothetical prote
41	192.5	3.6	1300	2 I53799	Gcl protein - huma
42	192	3.6	986	2 S49394	HsdR protein - My
43	192	3.6	1356	2 S12763	kinectin-1 - human
44	191.5	3.6	943	2 F65543	ATP-dependent RNA
45	191	3.6	1922	2 T00637	hypothetical prote

ALIGNMENTS

RESULT 1
T32534
hypothetical protein F15B10.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T32534
R:Wohlmann, P.; Murray, J.
submitted to the EMBL Data Library, December 1997
A:Description: The sequence of C. elegans cosmid F15B10.
A:Reference number: 221186
A:Accession: T32534
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1037 <WOH>
A:Cross-references: EMBL:AF036696; PDB:AA88350.1; GSPDB:GN00022; CESP:F15B10.2
A:Experimental source: strain Bristol N2; clone G15B10
C:Genetics:
A:Gene: CESP:F15B10.2
A:Map position: 4
A:Introns: 30/3; 89/1; 111/2; 218/1; 258/2; 288/1; 343/3; 370/4; 400/2; 443/1; 520/1;

Query Match
Best Local Similarity 12.9% Score 687; DB 2; Length 1037;
Matches 251; Conservative 169; Mismatches 408; Indels 174; Gaps 35;

QY	101	DLPSPFENAHDE--YLOLNLLOPTLVNKLVRDYLDCMEEE-----	142
DB	63	DLKNSILSNADDERLYKDIINTYLO--TYLPKCTVHKLLN--CSNREYKLSDFHY (LID)FEGF	120
QY	143	LITIEDRNRIAAENNGNSGVALLKRIYOK---ENMTSALN--VLEQTG--NNELYQ	195
DB	121	LRETFEKKVYLAIDSYPOYIDAVALRKELEREEDNODSFKKLITLTVLIGCAVAY	180
QY	196	EL--TSSDCSESNALEINLSQVNGPOVEPOLISTVQPLNEKVMEN--	243
DB	181	DIWYTTSEKSSNNLVE-----AKOFIA-----KVLKLNKGKREYQIIN	221
QY	244	SSESPADSSVSESPTSLSAEGSVSCLDSLAGNS-----NMGSDCTMGSDNLE--ENV	296
DB	222	ASRQRLNGRIYICPVHESATEMMVYLGTAALNFRMIRINVDNIQVNSIIRCVIESV	281
QY	297	AAASPEPELOLPYQMEVAQAPLEKNNIILCPTSGGTRAVYIAKHILKKKKASEP	356
DB	282	RORIHQRORCLRNVOELQVALQKNTIVAPTSQKTVIAANLIKHFPSRSEGR	341
QY	357	GKVIYIV--NKVLVEOLFREFQPFLLKMYRVIGLSGDTQKISPEVVKSLIISTAV	415
DB	342	FKALFPMPSMILNQQ--AAISISSYLDHYHTQIIGSD--NVPNTVYLSQKLLIVATQ	397
QY	416	ILENSLILNENGDAVQ-----LSDFSLLIIDECHHTKKEAVYNNIMKHYI	470
DB	398	MIYNLCNHRHNSLDDSRILDDQFLSTFTIIFDECHNTVKNKSPYSIMREY--HYLKNGG	455


```

Db 21 LESIYADPEKGESELELPLGKIDELKTHS-ENSEFESKOLYODLKSNI:SIADP-----E 75
129 LLYRDL-----DKMEELITIEDRRKIAAENNGESVRELKRIYOKENFSAF 181
Db 76 RLYKDIIMAYLGSNFPKILHLYNLSGRN-VKLSERYLLIDHLEGLRLIFPK-----VY 129
OY 182 LNVLRGTGN-NELVDELTSDESNAEJENLSQVNGPOVEBOLLSTTVLHLEKFWJH 240
Db 130 LNVLSYPOYSFAVKLL-KDELDRHEEDNN-----DPEFIKLIIRTYLIGEDVAVY 182
OY 241 ENNSSESSPADSSVSESPTSUA-EGSVASCLDESLSGHSNMSGDSQTMGSDSESNWAAR 299
Db 183 MYTIMKAMQTSIIINIRVDSIAPENSASRL-----VISVRCR 221
OY 300 ASPEPELOLRPYOMEAQALLEGKNITICLPFGSGKTRVAVYIADHLKKKAKASIPGV 359
Db 222 IHQRFQCLNRYOELCOVALQOKNTIYAPFGSCRTVIAANIIEHPESK---SSEO-- 276
OY 360 IYLVNVLVLEQLFRKEFQPLKKYRVYIGSGDTQKLSPEFVVKSCDIIISTAOILEN 419
Db 277 -----KREKTIQOSD--NVPTRVYQSKDLYATPOMIYN 311
OY 420 SLNLNGEDAGVQ-----LSDPSLIIDECNHTNKEAVYNNIMRAYIHOGLKNNRK 472
Db 312 -LCN-EHRDPLDEYEPPEQFLSTFIIFPEDECHNTVKNSSPVNREV-HYIKNMG 367
OY 473 KKKRPVILPOLLGLTASGVCATQAKAEHILKLCNLDATFKYIKINLDOLKNI 532
Db 368 PECH---SPFOITIGLTAIGTGKMKCMQVRSYIACLCANMPOVKELSIKONLELDHN 424
OY 533 QERCKFATADATREDP-----FKKLEITMTRIQYTCOMSPMSDFGTP-----Y 578
Db 425 PPTDQVSCGERSNDPIEMFTKRLKQMOQVEEDLRTILKNEP YKYEIIPDCKHNY 484
OY 579 EDWALQMEKKAANKGNRKEVCALHRLKYNALQINDITRMIDATYHLEFYNEEKKE 638
Db 485 ENNISQNRQCVSLAGSRNKLTLIEVL-----DVLKQVYMFQOMKIFNNYFKRYP 534
OY 639 AV---IEDDDEGGDDCYCGDEDEDLKKPKLD-----ETDFLMLIFPENKRLK 688
Db 535 GKKRIFQDCCFYALSYNINFPNVALKRYLEKEIGPERIRNFTON--NKLIMDN--CHR 590
OY 689 RLAEPEYENKLTKLNTIMEQYRTESARGIIFTKROSAYALSONIIEKEKFAEYG 748
Db 591 ELVGTSAENPMIARTVQFILDONEQTSDFRALIFVRKKEEDFLNYVL--NDRILHEIG 647
OY 749 VKAHLIG---AGHSSEFKPMTONEQKEVYSKFTGKINLLATVAEEDLIDKECNVI 805
Db 648 IKSDMWSQOKKSTASSADISASKQOMEKLMKMFADGENQILVSTVAEEDLIDPECSVI 707
OY 806 RYGLVNTNEMIAVOARARADESTYVVAHSGSVIEH--EYVNDREKMKYKAIHCYON 863
Db 708 KYNYAMINETAHVORGRARARANSKCVLITNS---LALHVESNMLKAKENMLMETISLON 764
OY 864 MKPE-----EYVAKTILELOMOSIMEKMKTKRNIKHKNNKNSLTLTLCNCSVLACS 917
Db 765 SPEEFQCVDEESNKKYMPRIQREDTQKAOIRKQI-----NRNLYVKKIYMKCDIVLCIN 819
OY 918 EDIHVTEKMHVNMTPF-----KELYIVENKALQKKCADYQINGELIC--KCGQAWGT 970
Db 820 KDIR-SKNQYIYVNCNGFMSLVARIPLEQRASNK-----FNSTGSEICIGEGSGSKIG 874
OY 971 MMYHKLDPCLKIRNPVYVFFKNSNKKYKQKAVEL-----PITPNNL 1014
Db 875 LIDVNTVNLPCLKVKKISLLIESTNRIIVKQMKNILDHFHTTLCORL 924

```

RESULT 4

T46312
 hypothetical protein DKFZP434J111.1 - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000

C:Accession: T46312
 R:Ducresthoefft, A.; Lauber, J.; Mewes, H.W.; Weill, B.; Milmann, S.
 submitted to the Protein Sequence Database, January 2000
 A:Reference number: Z23036
 A:Accession: T46312
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-398 <AAA>
 A:Cross-references: EMBL:AL137608
 A:Experimental source: adult testis; clone DKFZP434J111
 C:Genetics:
 A:Note: DKFZP434J111.1

Query Match 9.7%; Score 516; DB 2; Length 398;

Best Local Similarity 33.3%; Pred. No. 2.2e-19;

Matches 141; Conservative 68; Mismatches 157; Indels 5; Gaps 14;

```

OY 594 MKREHYC-----AEHLKRYNEALQINDITRMIDATYHLEFYNEEKDKKRAVIETQNSDE 647
Db 2 DESRICALFLYTSHLKRYNDALIISEHARKKALDYLKDFSSVRAAGF----- 52
OY 648 GGDDEYCGDDEDEDLKKPKLDETDRLMTLFFENNMKRLAENPEYENKLTIN 707
Db 53 -----DETEODL-----TOR-----FEKLELSESVSHOPSNEPKLEDI 90
OY 708 IMEYRTRESARGIIFTKROSAYALSONITENEK--FAEYGVKANHILIGAGHSSEK 765
Db 91 LOEBETHLNDEYIT-ILFKTRALVDALKNMIEGNKLSFLKCI---LTGKTKNLNTG 145
OY 766 MTQNSQKEVYSKFR-TGKINLLATVAEEDIDKECNIVIRYGLVNTNEMIAVAKAR 824
Db 146 MTLPAQKCLIDAFKASGQHNILDIATSVADGIDIQCNLYLIEYGVNVIKMIYIKCKGR 205
OY 825 ADESTYVVAHSGSVIEHETVNDREKMKYKAIHCYONMKPEEPAHKLIELQMSIEK 884
Db 206 AAGSCFELT-SNAGVIEKQIMNYKEMKNNDSILRLQTWDEAVYREKILMLQ14--EK 261
OY 885 KKKTRNIAKH--KNNSPLITFLCNCSVLACSGEDIHVTEKMHVNMTPFKEIVR 942
Db 262 FTRDQERKKRPDKENKLL--LCRKCALACTADYKAVIECEYITVLGDAPKQYVSR 318
OY 943 ENKALQKKCADYQINGELIC--KCGQAWGTMMVHKGDLPLKIRNPVYVFNNSIKQ 999
Db 319 PPKP-PKQSSPEKRAKIFCARNQNSHDMGIVHKYKTEIPVLIKIESVVEIATVYL 377
OY 1000 YKKW 1003
Db 378 YSKW 381

```

RESULT 5

T20332
 hypothetical protein D2005.5 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T20332

R:Wilkinson, J.
 submitted to the EMBL Data Library, September 1996

A:Reference number: Z19258
 A:Accession: T20332

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
 A:Residues: 1-1134 <WLL>

A:Cross-references: EMBL:Z79752; PIDN:CA802082.1; GSPDB:GN00019; (PSE) D2005.5

C:Genetics:
 A:Experimental source: clone D2005

A:Map position: 1
 A:Introns: 33/2; 91/2; 116/3; 144/1; 319/1; 421/3; 482/2; 511/1; 684/3; 757/1;

Query Match 9.4%; Score 499; DB 2; Length 1134;
 Best Local Similarity 23.5%; Pred. No. 7.3e-18;
 Matches 193; Conservative 134; Mismatches 288; Indels 20; Gaps 29;

A: Experimental source: strain Otsay
C: Genetics:
A: Gene: PAB0190
C: Superfamily: ATP-dependent RNA helicase e1f-4A
P: 33-40/Region: nucleotide-binding motif A (P-loop) #status: active
P: 131-136/Region: nucleotide-binding motif B
P: 135-138/Region: DEAD/H motif

Query March 8.5% Score 453; DB 2: Length 752;
Best Local Similarity 25.3%; Pred. No. 9,4e-16;
Matches 183; Conservative 124; Mismatches 259; Indels 156; Gaps 28;

QY 306 LQLRPPQEVAPALEGNNIIICLPTGSGKTRAVYIAKDHLKKKKADEPKYIVY-N 364
DB 11 LQPRLYQ-ELIYAKCKENKNCILVPTGLGKTIAMADIYRLNK-----YGGKVIAMAPT 64
QY 365 KVLIVE-QLRKKEQPLPKWRYVIGSGDTOLKISPEPVYKS---CDIITSTVLEN 419
DB 65 KPLVLOHAEETKFKFNNLPPEK---IVALTGE---VSPSEYKANNRAKVIILAIHQIVEN 117
QY 420 SILNLENGEDAGVQSDSLIIDECCHTNKEAVYNNIMRHLMOKLKNRRKKKNNKI 479
DB 118 DLT-----GKINLEDVYLFPDEAHAVGNAYVYIAREYLNDAKN----- 159
QY 480 PLPQILGLTASPGVGGATKQAAEHHILKLCANDAPTITKYENLDOLKQVQVYK 539
DB 160 -PHVIGLTASPG-----SSQEKILEVNLNIGIEHIEYRESENPQVKPVYQS- 1KF 207
QY 540 AIAADTRDPKE--KLEIMTRIOTYCOMSPMSDFQPRPEQMAIQEKKAARKNK-- 595
DB 208 EMIKVELPGLTKREVKLLRDMR---DSLKPLAAGL--TESASADLPKTIILKAIQII 261
QY 596 -KERVCAEH-LRK-----YNEALQINDTRIMDAY-----THLETFYNEEKIKKRVATED 643
DB 262 NEETAKGNHDLRKLFLPHAMALKLHAEILELDTQSLARVYLKKLYEAK--- 312
QY 644 DSDEGDDPEYCDGDEDEDLKKPLKIDETDRFLMILFPENNKMLKRLAENPEYVNHKLT 703
DB 313 -----TGSTRAKSELEFLDRRMKALALLQ-----AKELGILHPKMEV 350
QY 704 LNTIMEQYTRTEESARGIIFTKTRQSAVYALSQWITENEKFAEYGVKANHILGAHSEEF 763
DB 351 LKELVKEQLSK-KENSKILVETNRETARKVVEELTK-----GIKAKRVQCATKEND 403
QY 764 KPTQNEQKEVYSKRTKINLLATYAEGLDKCEKNIVIRGLVNTETAMVAKGR- 822
DB 404 RGMRSRQEKLLIDSFARSEFVYLVATVSGEGGLDQPEVLDVFEVPASAIRSVAKKRT 463
QY 823 ARADESTVLAHNSGVILEHETVNDPREKKMYAINHVV-----QMKPEYVAHKLE 875
DB 464 GRQKRGVYVILAAQSTRQREAVYVSSKKHKEQKQREITRNYSQAIKREKLSLENYKKITE 523
QY 876 LQMSIMKKKTKRNIAKHKNPSSLTFLCKNCSYLAGSGEDIVLEKNNHVMPTPEF 935
DB 524 KEMEREKEEKGEGREGKVVVDSRELRS-----EYVKRLKILVKEIV 567
QY 936 KEL-----YIVENKALOKKADYQING-----ELICSGQAWGIMM 972
DB 568 KTLQGVDTIISDEVAIEKKSANDPQSIIDGRLEPDQVARKLQSYPRPVYIVESGLYGRN 627
QY 973 VH 974
DB 628 VH 629

RESULT 7
D71203
probable ATP-dependent RNA helicase - Pyrococcus horikoshii
C: Species: Pyrococcus horikoshii
C: Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_chapter 10 May 2001
C: Accession: D71203
M.: Ohnuki, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kishida, N.; Ogu

DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic *A. Reference number: A71000; MUID:98344137; PMID:9571194*

A:Accession: D71203

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-650 <KAW>

A:Cross-references: GB:AP000007; NID:93236134; PIDN:MAA1011; PID:94258336

A:Experimental source: strain O73

A:Note: this accession replaces an interim accession for a sequence replaced by GenBank

C:Genetics:

A:Gene: PH1896

C:Superfamily: ATP-dependent RNA helicase eIF-4A

C:Keywords: ATP; nucleotide binding; P-loop

F:133-40/Region: nucleotide-binding motif A (P-loop)

F:131-136/Region: nucleotide-binding motif B

F:135-138/Region: DEAD/H motif

Query Match 8.3%; Score 441.5; DB 2; Length 650;

Best Local Similarity 25.9%; Pred. No. 36-15;

Matches 185; Conservative 122; Mismatches 26; Indels 147; Gaps 29;

306 LQRPYQVQVADPQALGKNTITCLPTGSKTRVAVYIADKH. KKKKASHGKAVYLV N 364
 11 LQRPYQVQVADPQALGKNTITCLPTGSKTRVAVYIADKH. KKKKASHGKAVYLV N 364
 365 KVLVE--QLFRKQRPFLKMYRVIGSGDTQKISPEVYVKS---CDIISTAGILEN 419
 65 KPLVLDHAEFRKFTFLPREK---IVALTGE---ISPM RVAMARAVVATPOTIEN 117
 420 SLINLENGEDAGVQSLIDFSLIIDECHTNKCAVYNNIM-HYLMQKLNKKKKENKPV 479
 118 DLV-----VGRSLDEVLIIIDFENHRAVGNVAVYIAPYIKO-AKK----- 159
 480 PLDPOLIGTASPVGATQAKAEHILKLCANLDAFTLYKENDOLKNOIOEPCKKF 539
 160 --PHVIGLTAISP-----STPEKIMEVLRMLGIEHI YNSNSPDVKKVYQV--TRF 207
 540 AINDADREDPEK--KLIEIMTRIOYCOMPSMDFGTQYEDMALOMKKAKKKNR-- 595
 208 EAWKVELPELYKVEKRLKRLDMR---DSKPLAEAGL--LDSTSDIPKREILRAQIT 261
 596 -KERVCAEH-LRK-----YNEALOINDTIRIMDAY-----THLETYNEKDKKFAVID 643
 262 NEMKAGNDLRLKLLFLHMAKLHAIETLEQSLALHAYIKRLYEAK----- 312
 644 DSDGDEDEYCDGDEDEDLKKPLKIDETDRFLMTPFNKMKLRIANPEYENKLIK 703
 313 -----ACSTRASKELFLDKRKKKALALLIO-----AKELGIDHKKMEY 350
 704 LRMTIMEQYTRTESARGITFTKROSAYALSONITENKFAVGVKAKHLIAGISSEF 763
 351 LKGLIRBQRLR-KENSKIIVFTYRTAKRV-----EKLKMGIRKARFVQASREND 403
 764 KPTNOROKEVISKFRITGLINLIATTVAEGLDKECNIVIKYGLVTN-VAWQAKGF- 822
 404 RGSQSRQKILIDAFARSEPNNVATSVGEGIDVPEVDLIVTEYEPSS-KSVQKRCPT 463
 823 ARADESTYLVVAHSGSVLEHETVNDFERKMYKATHCV-NMKPEEVANLDELQMSIM 882
 464 GRCQPCGVVILMAQGRDEAVYSSROKRIKIMETIRNVSQ-----VRHRLQSLIESV 518
 883 EKKMKTKRNTAKKKNPSSILITFLCKNSVLAACSGEDINVEKMHVNNIPEKEL----- 938
 519 KREVNE---IKKKETGKIVKIDSRKLS-----EVKKLKLTLGIRIEVRLTEVND 569
 939 YIIRKNAKOKKACADYQING-----ELCKGQAWGMAVH 974
 570 YIYSEDAIERSANDFOSIIDGRLPDQYKRLKEAVPRVITIEGOLGIRNHH 624

RESULT 8
 A69432

ATP-dependent RNA helicase homolog - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 4-May-2001

C:Accession: A69432

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.F.; Kelchum, K.A.; Dod

: Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.C.; Hill, S.; Kirsch, E

: Gilek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

: Nature 390, 364-370, 1997

A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artlich, P.; Kaine, B.P.; Sykes,

: Smith, H.O.; Wiese, C.R.; Venter, J.C.

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch

A:Reference number: A69250; MUID:98049343; PMID:9389475

A:Accession: A69432

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-741 <KLE>

A:Cross-references: GB:AE001002; GB:AE000782; NID:92689325; PIDN:MAH8786.1; PID:9264

C:Superfamily: ATP-dependent RNA helicase eIF-4A

C:Keywords: ATP; nucleotide binding; P-loop

F:132-43/Region: nucleotide-binding motif A (P-loop)

F:132-137/Region: nucleotide-binding motif B

F:136-139/Region: DEAD/H motif

Query Match 7.9%; Score 417; DB 1; Length 741;

Best Local Similarity 23.8%; Pred. No. 6-5e-14;

Matches 169; Conservative 114; Mismatches 230; Indels 36; Gaps 30;

303 EPELDQRPYQVQVADPQALGKNTITCLPTGSKTRVAVYIADKH. KKKKASHGKAVYLV N 362
 11 EPELDQRPYQVQVADPQALGKNTITCLPTGSKTRVAVYIADKH. KKKKASHGKAVYLV N 362
 363 KVLVE--QLFRKQRPFLKMYRVIGSGDTQKISPEVYVKS---CDIISTAGILEN 419
 65 KPLVLDHAEFRKFTFLPREK---IVALTGE---ISPM RVAMARAVVATPOTIEN 117
 420 SLINLENGEDAGVQSLIDFSLIIDECHTNKCAVYNNIM-HYLMQKLNKKKKENKPV 479
 118 DLV-----VGRSLDEVLIIIDFENHRAVGNVAVYIAPYIKO-AKK----- 159
 480 PLDPOLIGTASPVGATQAKAEHILKLCANLDAFTLYKENDOLKNOIOEPCKKF 539
 160 --PHVIGLTAISP-----STPEKIMEVLRMLGIEHI YNSNSPDVKKVYQV--TRF 207
 540 AINDADREDPEK--KLIEIMTRIOYCOMPSMDFGTQYEDMALOMKKAKKKNR-- 595
 208 EAWKVELPELYKVEKRLKRLDMR---DSKPLAEAGL--LDSTSDIPKREILRAQIT 261
 596 -KERVCAEH-LRK-----YNEALOINDTIRIMDAY-----THLETYNEKDKKFAVID 643
 262 NEMKAGNDLRLKLLFLHMAKLHAIETLEQSLALHAYIKRLYEAK----- 312
 644 DSDGDEDEYCDGDEDEDLKKPLKIDETDRFLMTPFNKMKLRIANPEYENKLIK 703
 313 -----ACSTRASKELFLDKRKKKALALLIO-----AKELGIDHKKMEY 350
 704 LRMTIMEQYTRTESARGITFTKROSAYALSONITENKFAVGVKAKHLIAGISSEF 763
 351 LKGLIRBQRLR-KENSKIIVFTYRTAKRV-----EKLKMGIRKARFVQASREND 403
 764 KPTNOROKEVISKFRITGLINLIATTVAEGLDKECNIVIKYGLVTN-VAWQAKGF- 822
 404 RGSQSRQKILIDAFARSEPNNVATSVGEGIDVPEVDLIVTEYEPSS-KSVQKRCPT 463
 823 ARADESTYLVVAHSGSVLEHETVNDFERKMYKATHCV-NMKPEEVANLDELQMSIM 882
 464 GRCQPCGVVILMAQGRDEAVYSSROKRIKIMETIRNVSQ-----VRHRLQSLIESV 518
 883 EKKMKTKRNTAKKKNPSSILITFLCKNSVLAACSGEDINVEKMHVNNIPEKEL----- 938
 519 KREVNE---IKKKETGKIVKIDSRKLS-----EVKKLKLTLGIRIEVRLTEVND 569
 939 YIIRKNAKOKKACADYQING-----ELCKGQAWGMAVH 974
 570 YIYSEDAIERSANDFOSIIDGRLPDQYKRLKEAVPRVITIEGOLGIRNHH 624
 914 ACSGEDIHVEKMHVNNIPEKEL---YIIRKNAKOKKACADYQINGEI 960

Db 220 PNEFKALK - LINEALKERLKLNDAGVINSIADVTTELIELNKKLFY----- 268
 Oy 574 GTOPYQOMAIQMEKKAAGKGNREKRYCAERLKYNEALINDITRMIDA-----YIHDE 627
 Db 269 -----DEEYKTELIVCSALMLMAKLELESQKSVPLNTN 306
 Oy 628 TTYNEEKDKFAVIEDDSDEGDEYCDDEDEDLKPLKLDETDRFLYLFENNKKM 687
 Db 307 KLSMOTKSAKSIYNDE-----KREAANLILKSVHEPKL- 342
 Oy 688 KRLAENPEYENKELTKRLNTIMEQYTRTEESARGIIFTKROSAYALSQYIENEFKFAV 747
 Db 343 -----KGVDMVNNILEK-----NKDERIIFQAYROTVEKIVNLTQN----- 381
 Oy 748 GYAKHLILGAGHSSEKPMQONOKREVISKFRGKINLLIATVAEGILKECNIVRY 807
 Db 382 GKRAIFPQCA-NKEGKSGOKQDAIEAFK-EGSVLAVTSVSEGITPSVNTILTY 439
 Oy 808 GLVTNEIAMVOAGRARADE--STYLVAHSGSGVIEHTVNDFREKMMYKAHCVQNM- 864
 Db 440 EAPVSEIRFIQRRGRAMRGEGKVYLLAKGTADNAVYSAL-YKERERKRL- KMG 495
 Oy 865 -----KPEYAHKTELQMOSTMEKMKTKRNIAHKK-----NNSTITFLKN 909
 Db 496 YLLNKLQKKFEKSEKEIEKEIEKEIEKESKTAVKEETKEEKEKTKIITLDFIKQ 555
 Oy 910 GSV-LACGED-----IHYEKMHVNMTP-----EFKPL-----YIVKE 943
 Db 556 IEKERSKEEDIKOEIKIPKKPIKIVDVRKNNAKLLHNYANLEKILEVDYVLSID 615
 Oy 944 NKALOKKADYOINGEI 960
 Db 616 RYVVERKTAEDFVNSII 632

RESULT 11
 S44849
 K12H4.8 protein - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 14-Sep-1994 #sequence_revision 12-May-1995 #text_change 30-Jun-2001
 C:Accession: S44849
 R:Favella, A.D.
 Submitted to the EMBL Data Library, May 1993
 A:Description: Sequence of the C. elegans cosmid K12H4.
 A:Reference number: S44851
 A:Accession: S44849
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1822 <FAV>
 A:Cross-references: EMBL:L14331; NID:g289702; PIRN:AAA2101...; PIR:g289703
 C:Genetics:
 A:Insertions: 17/3; 72/1; 116/3; 170/1; 216/3; 340/3; 401/3; 413/3; 512/2; 634/3; 694/3; 74
 C:Keywords: ATP; nucleotide binding; P-loop
 F:33-40/Region: nucleotide-binding motif A (P-loop)
 F:141-146/Region: nucleotide-binding motif B
 F:145-148/Region: DEXH motif

Query Match 6.6%; Score 350; DB 2; Length 1822;
 Best Local Similarity 23.5%; Pred. No. 6,5e-10;
 Matches 175; Conservative 118; Mismatches 249; Indels 202; Gaps 37;

Oy 309 RPYOMEVAPALEGKNIICLPTGSGKTRVAVYIAKDI-----IDKKKASEPKKIV 361
 Db 14 RDVQVELDKATK-KNIVQLGSGSGKTRVAVYIAKDI-----RAPP 67
 Oy 362 LVNKKVLLVEOLFKEKQPLKKVRYVGLSGDT-----OLKISFVYKSTDIITISACI 416
 Db 68 VVEKVNIVEDQ--QATHIEHTSPKQGVAGQTSGLMDSEKCEQKHHVAVITAGC 124
 Oy 417 LENSLLMENGEDAGVQLSDPSLIIDECHNT-NKEAVYNNIMRHYLMOKLKNRLAKEN 475
 Db 125 L-----LDLRHAYLKEIDEMCVLIDFECHHALGSHPRYSIMVY-----KLLKRD 170

Oy 476 KEVILPOLIIGTASPGVGATQAKASEHILKLCANLDAFTIKTKVENLIQIKNOLEP 535
 Db 171 KVV-----PRYLGITAGL-IAKAVAPEKLMQKLKLESANDS-VIETASD-LVS-SKVGAKP 224
 Oy 536 -----CKKFAIA--DATREDPFKKLEIMTRIOTYQOMSPMSDFQO----- 576
 Db 225 YEVVILCKKFELGCGIGNFDVIEIFDETAVFVNTTEFFRDLDDLDRPKIKESLKITR 284
 Oy 577 -----PYEQALQMEKKAAGKGNREKRYCAERLKYNEALINDITRMIDA-YIHDE 627
 Db 285 AVFROLGPMAN-----RTAOMWEKELGKILIKSOVLDPDKTLRIIMAKITSI 331
 Oy 628 TTYNEEKDKFAVIEDDSDEGDEYCDDEDEDLKPLKLDETDRFLYLFENNKKM 687
 Db 332 TI-----KRLI-----EPKMKIKISLEALRPVQGVIAFL 365
 Oy 688 KRLAENPEYENKELTKRLNTIMEQYTRTEESARGIIFTKROSAYALSQYIENEFKFAV 740
 Db 366 ETF--NPEFOKERM-KLEKA-----EHLSAIIFVDQRYAVSLMLMHKISW--- 410
 Oy 741 NKKFAVGYKAHHLGAG-----HSSEKPMQONOKREVISKFRGKINLLIATVAEFL 796
 Db 411 EKKFK--FVNPDYVAGASRLASSDSQGL-HKROTEVLRPRHNEINCLIAVSLEESV 467
 Oy 797 DIKECNIVRYGLVNTNEMIAMOAGRARADESTYLVAHSG-----SGVIE 842
 Db 468 DYKQCNLYKPRDIDMSRYVOSKGRARRAGRYIYIEKRTAAYCSKLPSDITRIV 527
 Oy 843 HETVNDFREKMMYKAHCVQNMKP--EYVAHKTELQMOSTMEKMKTKRNIAHKKNP 900
 Db 528 HNOIIPIEENGYTK--YCAELLPLINSPIKHA1-----VLKNPMNKI----- 569
 Oy 901 SLITFLCKKCSYLACSGE-DIHVIEK-----MHVNMTP-----EKE 937
 Db 570 AQMAVALEACROLHLEGELDDNLLPGRSIAKLEHIDEDDEYAPGIAAVYSSKKRQ 629
 Oy 938 LYIVRENKAL-----OKKADY 954
 Db 630 LYDKIARALNSESFEADKECFIY 653

RESULT 12
 E69055
 ATP-dependent RNA helicase, eIF-4A family - Methanobacterium thermoautotrophicum (str
 C:Species: Methanobacterium thermoautotrophicum
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-May-2001
 C:Accession: E69055
 R:Smith, D.R.; Doucette-Stamm, L.A.; Delonghery, C.; Lee, H.; Dubois, J.; Aldredge, T.
 J. Bacteriol. 179, 7135-7155, 1997
 A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu
 A:Reference number: A69000; MUID:98037514; PMID:9371463
 A:Accession: E69055
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-738 <MTB>
 A:Cross-references: GB:A600903; GB:A600666; NID:g2622514; PIRN:AA85892.1; PIR:g262
 A:Experimental source: strain Delta H
 C:Genetics:
 A:Superfamily: ATP-dependent RNA helicase eIF-4A
 C:Keywords: ATP; nucleotide binding; P-loop
 F:38-45/Region: nucleotide-binding motif A (P-loop)
 F:134-139/Region: nucleotide-binding motif B
 F:138-141/Region: DEXH motif

Query Match 6.4%; Score 342; DB 1; Length 738;
 Best Local Similarity 24.7%; Pred. No. 4.6e-10;
 Matches 172; Conservative 106; Mismatches 234; Indels 184; Gaps 31;

Oy 302 PEPEIDLRPYOMEVAPALEGKNIICLPTGSGKTRVAVYIAKDI-----IDKKKASEPKKIV 361

Db 12 PE-KIEARTYQOILAAADYIRKNSMIVAPTALAGTIVAVLAERLKKYMS---KVL 66
 QY 362 L-VKKVLLVLDLFRKEQOPFLKKRYVIGSGTOLKISF-PEVVK-----SCDIIISTAQ 415
 Db 67 LSPKPLAIOH--EESFRFM-----LATQSLIGS-IPKPKRKRMIKSOIISATPO 116
 QY 416 ILNLSNLNLENGEDAG-VQISDFSLIIDECHHTNKEAAYNNIMRHYLMOKLNNRKK 474
 Db 117 TVEEDIL-----AGHYDLRDVSLIVFDECHRAVGSYVFLASN-----IQMAR----- 161
 QY 475 NKPIPLPOLGLTASGCGATKQAARAEHILKCANL--DAFTIKT-----VKENDQ 527
 Db 162 -----HPLILGLTASGAD-----EDKIKTVCENLPMNEVAVTECHIDPVPYLPK 207
 QY 528 LKNOIOEPCKKFALADATREDPFREKLEIMTRIOTYCOMSPMSDGTOTYQOMAIQMEK 567
 Db 208 IKIEVVK-----VRRTPLEEDRELRLRVLKN-RKLMLKNLCVIL-----TISVCK 252
 QY 588 KAAKKG-----NRKERVCAEHLKRYNEALQINDTIRMIDAYTHLETFYNN-EKDKKEAVIE 642
 Db 253 KDLKAGRVONRIARSTSPRACYSIALSCINVEHLELET----- 298
 QY 643 DSDDEGDDDEYCDGDEDEDLKKPLKIDETDRPLMTLFFENMKLRKLAENPEY----- 696
 Db 299 -----OG-----IRPL-----HOYLLRLKEKKTKAAGLLAIHPTRAMHL 334
 QY 697 -----ENKLTUKLNTIMEOYTTEESARGIIFTKROSAYALSOIKTENEFAY 747
 Db 335 TRRAMSGVHPKLDRLME-TLRRELKGDG-AAIIVTQFRDLLELYO-----CKRE 386
 QY 748 GYVAHHLIGAGHSSEFKPTONEKEVYSKFTGKINLJATVAERGLIKKCNVIRY 807
 Db 387 GINAVKFGGNSRSGEGLQKQRODIKSFRCNNDVILSTVAEGITIPVLDLVAM 446
 QY 808 GLVTNELAMVQAGRAADESTVYLAHSGSGVIEHETVNDPREKKMYK?HCVQNNKPE 867
 Db 447 EPVPEELRMIOHRCGRGKRGRMYV-----LITEKTRD---EAYYSSTIKERSMK-F 496
 QY 868 EYAHKILQMOSEMKMKTKINKIAKHNNPSLITFUKNSVLAAGSNDIHVEKKN 927
 Db 497 NLGGSNVNVNPIMER-----SDEGPFYIADSK 525
 QY 928 HVN-MTPFEKEDYIVRENKALOKKADYOINGEII 961
 Db 526 EVNSRYLRELKIGVDELKPL--AVGDYQISDITI 574

RESULT 13

ATP-dependent RNA helicase homolog eIF-4A [Imported] - Halobacterium sp. NRC-1
 C:Species: Halobacterium sp. NRC-1
 C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 10-May-2001
 C:Accession: D84386
 R:Jung, W.V.; Kennedy, S.P.; Mahapatra, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leitman, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Madlocks, D.G.; Jablo, J.; Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Edhardt, H.; Lowe, T.M.; Li, A.; Title: Genome sequence of Halobacterium species NRC-1.
 A:Reference number: A84160; MUID:20504483; PMID:11016950
 A:Accession: D84386
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-784 <STO>
 A:Cross-references: GB:AE004437; NID:g10581766; PIDN:AG20456.1; GSPD8:GN00138
 C:Genetics:
 A:Gene: eif4a
 C:Superfamily: ATP-dependent RNA helicase eIF-4A

Query Match 5.9% Score 313.5 DB 4 Length 784.
 Best Local Similarity 21.6% Pred. No. 1.5e-08.
 Matches 126; Conservative 91; Mismatches 190; Indels 175; Gaps 22.

QY 306 LOIRPYOMEYAPALGKNNIICLPYSGSKTRAVYIAADHLDKKKKASEPKVIVYVVK 365
 Db 18 IEARVYLODLAAAROG-HTVCLPTGLTKTVSLVTVYRL---ADDAQATALIAPT 72
 QY 366 VLLVEOLFREFQOPFLKKWR-----VIGSGDTOLKISFPEVKS---CIIIIS 412
 Db 73 KPLVQ-----HAGTYRELAIPDDDVYVFTETR-----PDKRAANTDAKAVYA 118
 QY 413 TAOILNLSNLNLENGEDAGVQISDFSLIIDECHHTNKEAAYNNIMRHYLMOKLNNRKL 472
 Db 119 TPQVYENDLVG-----GRIDMDVYCHTFECHRAVGATVYVAERYHAD----- 164
 QY 473 KENKPYPLPOLGLTASGCGATKQAARAEHILKCANLDAFTIKTKYKENDIQKQI 532
 Db 165 -----AAPLVYAMSAPG-----GTEAEIRTCENLGVNVEVTEI----- 202
 QY 533 QEPCKFALADATREDPFREKLEIMTRIOTYCOMSPMSDGTOTPE---QVAVLMEKKA 589
 Db 203 -----DA-----DVGERTHOTDVQV----- 217
 QY 590 AKKNRERVCAEHLKRYNEALQINDTIRM--DAYTHLET-----YNEE 633
 Db 218 -----EYVTLR-----EILEVRDAINDVIEDRLAKREIGVTKASSPDISOKIMEI 265
 QY 634 KDKKFAVIEDDSDGDDDEYCDGDEDEDLKKPLKIDETDR-LMTLFFEN----- 683
 Db 266 KARLOOLIDDSDGOGNSVAEYVK--LKRVELVETQSVESVRRFERKANAANIAG 323
 QY 684 -NMMLKRLAENPEYENKLT-----KLNNT--INBOYRTESAGGIIIRKQSA 731
 Db 324 ASASRLVSEFPAVRAMTAREFDGLHRKFKQAMLLAETIGIDGDRIVITVTSKITA 383
 QY 732 VALSOMITNEKFAEYGVKAHHLIGAGHSSEFKPTONEKEVYSKFTGKINLJATIV 791
 Db 384 EALTAFLEIGH-----FDRRFVCGGDGSDGMDTQTEQRETLAEFRNGDEVLVISTV 436
 QY 792 AEEGLDKECNIVIRYGLVTNELAMVQAGRAADESTVYLA 833
 Db 437 AEEGLDPEVDLVLFEPVPTAIRSVQKRGKROTAGRNVV 478

RESULT 14

S48436
 Probable RNA helicase YIR002c - yeast (Saccharomyces cerevisiae)
 M:Alternate names: hypothetical protein Y1B2c
 C:Species: Saccharomyces cerevisiae
 C>Date: 02-Dec-1994 #sequence_revision 02-Dec-1994 #text_change 19-Apr-2002
 C:Accession: S48436; S50885
 R:Badcock, K.; Churcher, C. submitted to the EMBL Data Library, August 1994
 A:Reference number: S48432
 A:Accession: S48436
 A:Molecule type: DNA
 A:Residues: 1-993 <BAD>
 A:Cross-references: GB:Z47047; EMBL:Z8062; NID:g603997; PID:g764447; MIPS:YIR002c
 R:Voss, H.; Tanamas, J.; Teodoru, C.; Valencia, A.; Sensen, C.; Wiemann, S.; Schwaiger, Vast 11, 61-78, 1995
 A:Title: Nucleotide sequence and analysis of the centromeric region of yeast chromosome
 A:Reference number: S50795; MUID:95282515; PMID:7762303
 A:Accession: S50885
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-993 <VOS>
 A:Cross-references: EMBL:X79743
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1994
 C:Genetics:
 A:Gene: SGD:MPI1
 A:Cross-references: SGD:S0001441
 A:Map position: 9R
 C:Keywords: ATP; nucleotide binding; P-loop
 F:107-114/Region: nucleotide-binding motif A (P-loop)
 F:205-210/Region: nucleotide-binding motif B
 F:209-212/Region: DEAD/H motif

Query Match 5.7%; Score 303.5; Db 2; Length 993;
 Best Local Similarity 20.3%; Pred. No. 67e-18;
 Matches 143; Conservative 118; Mismatches 246; Indels 197; Gaps 25;

```

OY 307 QLRPYMEVAQPALEGNIIICLPTSGKTRAVYIAKHLDKKKASEGKVIYVNV 366
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 86 EVADYQYIVYKSL-FQNTLCALPTGMKGTFIASVLMVNFRTKKA---KILFADPR 140
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 367 LVEQLFRKFEQPLKKWYVIGLSGD-----TOLKISEPEVYKSCDII-STAOILNS 420
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 141 PLVAQO-----IKACLGITGIPSDQTAILLDKSRKNKEEIMANKRVFATPQVEND 192
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 421 LVLNLENGEDAGVOLSPFSIIIDECHHTNKEAVYNNIMRYIMOKIKNNIKENKPVIP 480
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 193 L-----KRGVLDPKDIOVCIVIDEAHRAGSSAYTNVAF-----IDRENSSY-- 234
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 481 LPQILGLTASPGVGATKQAKAEHILKCANLDAFTIK-TYKENDOLKNQIOECPKRP 539
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 235 --RLALATAP-----ASDLEGOEVYNNIDISKIEIRTEESMDIVYMKKKKKKI 284
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 540 AIAADATREDPFKEKLEIMTRIQTQCMSPMSDFGTQPEQWALQM-----EKKA 590
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 285 EV-----PLLEIEIDIE-----QLGMAYKPVLOQALFEGIVLDDPSQINAF 327
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 591 KKGNRKRVCA-----EHLKKYNEAL-----QINDIRMDAYTHLETYNEEKDKKAV 640
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 328 KAMQOSOKITANPTIEGIMWNNPFIOLLNNYGMKIKRKITY-GIRTFIRYQNK --- 382
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 641 IEDDSGDEGDECDDEDDLPKLKIDETDFLMTLFFEN-NMLKLAEN-----PE 695
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 383 -----CTEETTYNLMK-----STNKIAAEFYHPLINIKAKCENYISDKR 424
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 696 YENE-KITKIRNTIMEQYTTESARGITFTKROSAVIALSOMI-----738
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 425 FVGHGKLOCVRDELMPFOGRGDSRVLIPTELRESALIVKEILSVADQIRPHIRIGQ 484
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 739 -----TENEFPAEVG-----YKANHLLGAGH 759
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 485 ARAKEGDEKYYTRKNAKPKRKVVERLHROEKEFLAEITTKRANDKLKSRRIGSSE 544
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 760 SSEPKPTONEQKEVISKFRGTQINLIATVAEGSLDKECNVIRYGLVINEIMAVCA 819
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 545 EAQISGNOMKQKEVILHNFKKGEVNVLYGTSIGEGIDIGEVLLICYDLSPTNIGR 604
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 820 RGR-ARADESTYVLAHSGSGVIEHETVNDFREKMYKALHCVQNNK-----PE- 867
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 605 MGRTRGRDKIVLIFSSNESYKFERAMEDYSTLCALISKQIDYKSDHIFEDYIPBC 664
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 868 -----EYAHKILLEOMOSIMKKKKTKNTAK 894
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 665 HETLITINDENELIEMEDEVDEVIATQCMGKKKVPKKAITK 708
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 15

T31323
 ATP-dependent RNA helicase, eIF-4A family - Cenarchaeum syntosum
 C/Species: Cenarchaeum syntosum
 C/Date: 11-Jan-2000 #sequence-revision 11-Jan-2000 #txcyl-change 11-Jan-2000
 C/Accession: T31323
 R/Schleper, C.; Delong, E.F.; Preston, C.M.; Feldman, R.A.; Wu, K.Y.; Swanson, R.V.
 J. Bacteriol. 180, 5003-5009, 1998
 A/Title: Genomic analysis reveals chromosomal variation in natural populations of the un
 A/Reference number: Z20994; MID:98422450; PMID:9748430
 A/Accession: T31323
 A/Status: preliminary; translated from GB/EMBL/DDHJ
 A/Molecule type: DNA
 A/Residues: 1-502 <SCH>
 A/Cross-references: EMBL:AF083072; NID:93599393; PID:935994-6; PIDN:ANC62714.1

Query Match 5.3%; Score 284; Db 2; Length 502;
 Best Local Similarity 24.3%; Pred. No. 2.6e-07;
 Matches 133; Conservative 94; Mismatches 207; Indels 114; Gaps 23;

```

OY 303 EP-ELQLRPYMEVAQPALEGNIIICLPTSGKTRAVYIAKHLDKKKASEGKVIY 461
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 12 EFGAVERRRYQVGLBOAIR-ENCTIVLPTGLKTAVALQVIAHVIDERCAIPLATIRV 70
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 362 LVNKKVLLVQLFRKFEQPLKKWYVIGLSGDTQ--KISPEVYKSC--DIIEIAVL 417
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 71 LVN-----QHRQPLG---RALTISDITVGTGETIPRRKKMGSGVIGAIPTIA 116
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 418 ENSLNLNLENGEDAGVOLSPFSIIIDECHHTNKEAVYNNIMRYIMOKIKNNIKENKPV 477
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 117 RMDI-----ERGLVPLBOFGVIVIDEAHRAGVAYSSIAF-----A 153
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 478 VLPQLGLTASPGVGATKQAKAEHILKCANLDAFTIKTYKENDOLKNQIOECPKRP 537
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 154 VQDNRSMVGMTATL-----PSEREKAE-----TGTLLSSIAQRTEDDPVAVVLIAT 205
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 538 KPAIADATREDPFKEKLEIMTRIQTQCMSPMS-DFGTQPEQWALQWKKAAKKNRK 596
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 206 EMTKYVLPPEMEIORLKLALD-ERYSSLKRCGYDGSNBSLSALRL-RHVVIGQNR 263
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 597 ERVCAEHLKRYEALQINDIRMDAY--THLETYNEEKDKKAVIEEDSIKSGIYYG 654
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 264 A-----AKPLTAIRITVYALNIFEAHGYTPFLKFCERTSKKKGVVAE----- 306
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 655 DQDEDEDLKKPLKIDETDFLMTLFFENKKMLKLAENPEYENKLTIKLNTIMEVYIF 714
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 307 -----LFEQDR-----NFGALIRAKAQAAGMEHPKPLETAV----- 341
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 715 TEESARG--IIFTKROSAVIALSOWITENEFPAEVGKANHLLIG-AGHSEI KPMIFQ 771
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 342 --RGARGKALVFTSVDSVDLI-----HSRLKAAGINSGLLIGAGE---KGLKKKK 389
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 772 KEVSKFRGTQINLIATVAEGSLDKECNVIRYGLVINEIMAVQNGR-ARALISTY 830
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 390 VETVAKFRGGDYAVLSTVVGEGDISVNVIVFDVNPSSIRYVORRGTRGKIAJRI 449
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 831 VLVAHSGS 838
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 450 IYLMAGKT 457
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: May 8, 2003, 15:55:12
 Job time : 62 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame-plus.p2n model

Run on: May 16, 2003, 10:02:54 ; Search time 65 Sec/nds
(without alignments)
4836.057 Million cell updates/sec

Title: US-09-515-363C-2
Perfect score: 5311
Sequence: 1 MSNCGSTDENFRYLISCFRA.....LPITPNLDVSTNCLFSDFI 1025

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 8827.4

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/cg2_1/uspro/us09515363/runat_16052003_100241_24711/app.query.fasta.1.1223
-DB=Issued_Patents.NA -OFMT=fastcap -SUFFIX=ini -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi
-LIST=45 -DOCCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09515363 @CGN.1.1.32.0runat.16052003_100241_24711 -NUPU=6 -ICPU=3
-NO_XIPY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELop=6 -DELExt=7

Database :

- 1: Issued_Patents.NA:*
- 2: /cg2_6/ptodata/1/ina/5A.COMB.seq:*
- 3: /cg2_6/ptodata/1/ina/5B.COMB.seq:*
- 4: /cg2_6/ptodata/1/ina/6A.COMB.seq:*
- 5: /cg2_6/ptodata/1/ina/6B.COMB.seq:*
- 6: /cg2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	472	8.9	301	1	US-08-143-576-7
2	335.5	6.3	7037	4	US-09-853-768-3
3	227.5	4.3	193	1	US-08-143-576-6
4	212	4.0	5852	4	US-09-853-768-10
5	207.5	3.9	6773	4	US-09-166-350-27
6	199.5	3.8	4868	5	US-08-139-937-12
7	199.5	3.8	4868	5	PCT-US93-11310-12
8	194.5	3.7	8789	1	US-08-328-254-5
9	191.5	3.6	10136	1	US-08-353-700-2
10	191.5	3.6	10136	5	PCT-US95-16215-2
11	188.5	3.5	4363	2	US-08-685-576-5
12	187.5	3.5	6921	4	US-09-643-597-117

13	187	3.5	5020	4	US-08-961-527-142	Sequence 142, App
14	186.5	3.5	5361	4	US-08-973-462-2	Sequence 2, Appl1
15	186.5	3.5	6152	1	US-08-973-462-1	Sequence 1, Appl1
16	180.5	3.4	3393	4	US-09-104-3248-1	Sequence 1, Appl1
17	180.5	3.4	3393	4	US-09-162-713-1	Sequence 1, Appl1
18	179	3.4	2001	4	US-09-134-001C-2628	Sequence 2628, Ap
19	174	3.3	5053	2	US-08-685-576-2	Sequence 2, Appl1
20	174	3.3	11958	4	US-09-134-246-8	Sequence 8, Appl1
21	173	3.3	3636	4	US-09-134-001C-1983	Sequence 1983, Ap
22	171.5	3.2	7100	4	US-09-308-375-1	Sequence 1, Appl1
23	167	3.1	11091	4	US-09-134-001C-2243	Sequence 2243, Ap
24	166.5	3.1	5661	4	US-08-938-105-2	Sequence 2, Appl1
25	164	3.1	4739	3	US-08-685-576-2	Sequence 1, Appl1
26	160	3.0	30549	4	US-09-134-001C-322	Sequence 322, App
27	159.5	3.0	2871	4	US-09-134-001C-1615	Sequence 1615, Ap
28	159	3.0	19307	3	US-08-836-022A-10	Sequence 10, Appl
29	159	3.0	19307	4	US-08-427-048A-10	Sequence 10, Appl
30	155.5	2.9	13425	4	US-08-961-527-151	Sequence 151, App
31	155	2.9	3825	4	US-09-208-742-3	Sequence 3, Appl1
32	155	2.9	4582	2	US-08-993-228-9	Sequence 9, Appl1
33	155	2.9	43676	3	US-09-356-952-12	Sequence 12, Appl
34	154.5	2.9	4248	3	US-08-678-614-1	Sequence 1, Appl1
35	154	2.9	7808	4	US-09-453-702B-247	Sequence 247, App
36	152	2.9	5093	1	US-08-468-036-23	Sequence 23, Appl
37	152	2.9	5093	2	US-08-376-843-23	Sequence 23, Appl
38	151	2.8	3741	4	US-09-541-782-9	Sequence 9, Appl1
39	151	2.8	3741	4	US-09-723-820-9	Sequence 9, Appl1
40	151	2.8	4308	4	US-09-592-054-1	Sequence 1, Appl1
41	150.5	2.8	3107	4	US-08-213-419B-1	Sequence 1, Appl1
42	150.5	2.8	8948	4	US-09-643-597-119	Sequence 119, App
43	148.5	2.8	371	4	US-09-222-575-46	Sequence 46, Appl
44	148.5	2.8	1891	4	US-08-973-462-3	Sequence 3, Appl1
45	148	2.8	3572	2	US-08-713-815A-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1
US-08-143-576-7
Sequence 7, Application US/08143576
Patent No. 5643761
GENERAL INFORMATION:
Applicant: Fisher, Paul B.
Applicant: Jiang, Hongping
TITLE OF INVENTION: METHOD FOR GENERATING A SUBTRACTED
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White, c/o Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/143,576
FILING DATE: 25-OCT-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 43563/JPW/ANC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 664-0525
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:

```

: LENGTH: 301 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHEetical: NO
: ANTI-SENSE: NO
US-08-143-576-7

Alignment Scores:
Pred. No.: 5,32e-45 Length: 301
Score: 472.00 Matches: 97
Percent Similarity: 97.988 Conservative: 0
Best Local Similarity: 97.988 Mismatches: 2
Query Match: 8,898 Indels: 1
DB: 1 Gaps: 0

US-09-515-363C-2 (1-1025) x US-08-143-576-7 (1-301)

OY 476 LysProValIleProLeuProGlnIleLeuGlyLeuThrAlaSerProGlyValGlyIly 495
DB 3 AACCGAGTATTCCTCTCTCAGATACAGGAGCTAACAGCTTCACCTGATGTTGACGCA 62
OY 496 AlaThrLysGlnAlaLysAlaGlnGluHisIleLeuLysLeuGlyAlaAsnLeuAspAla 515
DB 63 GCCACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 122
OY 516 PheThrIleLysThrValLysGlnAsnLeuAspGlnLeuLysAsnGlnIleGlnIlePro 535
DB 123 TTACATATTAAACCTGTTAAAGAAACCTGATCAACCTGATAAAACCAATACAGAGGACA- 181
OY 536 CysLysLysPheAlaIleAlaAspAlaThrArgGluAspProPheLysGlnLysLeuLeu 555
DB 182 TGCAGAAAGTTTCCCTGATGAGATGCACAGAGGAGGACGACGACGACGACGACGACGAC 241
OY 556 GluIleMetThrArgIleGlnThrTyArgGlnMetSerProMetSerAspPheGly 574
DB 242 GAAATTAATGACAGATTCACAACTTATGTCAATATGATGACATGTCAGATTTTGA 298

RESULT 2
US-09-853-768-3
: Sequence 3, Application US/09853768
: Patent No. 6444466
: GENERAL INFORMATION:
: APPLICANT: Donna T. Ward
: TITLE OF INVENTION: ANTISENSE MODULATION OF HELICASE-MOI EXPRESSION
: FILE REFERENCE: RTS-0217
: CURRENT APPLICATION NUMBER: US/09/853,768
: CURRENT FILING DATE: 2001-05-10
: NUMBER OF SEQ ID NOS: 91
: SEQ ID NO 3
: LENGTH: 7037
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (183)...(5957)
US-09-853-768-3

Alignment Scores:
Pred. No.: 6,03e-27 Length: 7037
Score: 335.50 Matches: 186
Percent Similarity: 36.468 Conservative: 113
Best Local Similarity: 22.688 Mismatches: 233
Query Match: 6,328 Indels: 288
DB: 4 Gaps: 43

US-09-515-363C-2 (1-1025) x US-09-853-768-3 (1-7037)

OY 309 ArgProTyGlnMetGluValAlaGlnProAlaLeuGluGlyLysAsnIleIleIleGly 328
DB 315 AGAAATATACAGTGAAGCTGCTTCAAGCAGGCTCTGAT--CATTAATATATGCTCTGT 371

```

```

OY 329 LeuProThrGlySerGlyLysThrArgValAla---ValTyrIleAlaLysAsp----- 345
DB 372 TTAACACTGGCTCAGGAGGACATTATGCTAGTACTCTACTAAAGCAGTGTCTC 431
OY 346 HisLeuAspLysLysLysAlaSerGluProGlyLys---ValIleValIleValAsn 364
DB 432 TATCTAGATCTAGGGAGCTTCAAGTAAATGAGAAAGACGCTTCTCTCTCAAC 491
OY 365 LysVal---LeuLeuValGlnGlnLeuPheArgLysGluPheGlnProPheLysLys 383
DB 492 TCTCAACACGAGTGTCTCAACAGCTGCAGCTGCAGACTCATTCAGATGACAG-- 548
OY 384 TrpTyrArgValIleGlyLeuSerGlyAspThrGlnLeuLysIleSerPhe----- 400
DB 549 -----GTTGGGAAATACCAAACTAGAAAGTAATGATCTTGACAAAGAG 596
OY 401 -----ProGluValValLysSerGlyAspIleIleIleSerThrAlaGlnIleLeu 417
DB 597 AGATGAAACCAAGAGATTACTAGACAC--CAGGTTCTCATTAATGCTGTAAGTCCG 653
OY 418 GluAsnSerLeuAsnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 437
DB 654 TTGAATGTTTAAAAAT-----GGTACTTATCACTGTCAGACATT 695
OY 438 SerLeuIleIleIleAspGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 457
DB 696 AACCTTTGGTGTGTTGATGAGTGATGATGATGATGATGATGATGATGATGATGATGAT 755
OY 458 MetArgHisTyrLeuMetGlnLysLeuLysAsnAsnArgLeuLysLysGluAsnYsPro 477
DB 756 ATGAGG-----CTCTGTGCA----- 770
OY 478 ValIlePro---LeuProGlnIleLeuGlnLeuThrAlaSerProGlyValGlyValAla 496
DB 771 ATTGTGCTATCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 827
OY 497 ThrLysGlnAlaLysAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 516
DB 828 TGGATCCAGAGATTGGAAGAAAGTTTCAGAACTA----- 866
OY 517 ThrIleLysThrValLysGlnAsnLeuAspGln-----LeuLysAsn 530
DB 867 ---GAGAAATTTCTTAAGAGTAATGCTGAACTGCAGACTGACCTGGTGTCTACAGAG 923
OY 531 GlnIleGlnGlnProCysLysLysPheAlaIleAlaAsp-----AlaThrArg 546
DB 924 TATCTTCTCAGCCATGTGAG-----ATTGTGTGATTTGTGACCATTTATTCAGAGA 977
OY 547 GluAspProPheLysGlnLysLeuLeuGlnIle----- 557
DB 978 AGTGGCTTTATGAAAGAGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1027
OY 558 -----MetThrArgIleGlnThrTyArgGlnMetSerProMet 570
DB 1038 TGTAAATATATCTGATCAATCAAAAGAAAGAGATTCATTAATTGCAAAACAAACTA 1097
OY 571 SerAsp-----PheGlyThrGlnProTyArgGlnIleValIle 583
DB 1098 TCAGACTGTGCGCGTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1157
OY 584 GlnMetGluLysLysAlaLysLysLysLysLysLysLysLysLysLysLysLysLysLys 597
DB 1158 ATGATGTAAGAGAGACTACAGAAATATACAAACATGAGCAAGAGAGCTGTCAAGAAA 1217
OY 598 -----ArgValCysAlaGlnHisLeu 604
DB 1218 TTTTATTTGTTACAGACACTTCTCTAAGAAATATACATGACATGATGAAAGAAATTC 1277
OY 605 ArgLysTyrAsn-----GluAlaLeuGlnIleValIleValIleValIleValIleVal 616
DB 1278 TCACCTGCTACTTACCTGAAATTTGTAAGCTCTTAAGTAATCAAACTGTCGAATTC 1337

```



```

QY 617 ILeArgMet1LeaSPaLatYrThrHis-----LeuGluTrpPheTyrAsn 631
   : : : : :
Db 1338 TTACGGAAATATTAACCAATATGAGCGACACACTTTTGAAGAGAGTATGATATATAT 1397
QY 632 GluGluLysAspLysLysPheAlaVal1LeGluAspAspSerAspGluTrpGlyAspAsp 651
   : : : : :
Db 1398 AGAAATACAGAAATATATGTGTGATG-----AGTATCTGTGATGATGAT 1445
QY 652 GluTyrCysAspGlyAspGluAspGluAspAspLeuLysLysProLeuLysLeuAspGlu 671
   : : : : :
Db 1446 GAG-----GATGAAGAAATATGAGAAAGAGAGAGAGCA----- 1478
QY 672 ThrAspArgPheLeuMetThrLeuPhePheGluAsnAsnLysMetLeuLysArgLeuAla 691
   : : : : :
Db 1478 ----- 1478
QY 692 GluAsnProGluTrpGluAsnGluLysLeuThrLysLeuArgAsnThrLysMetGluGlu 711
   : : : : :
Db 1479 -----GAGCAAAATTTTCTCTCTCTCTTTTCACACATTTTCTGTC----- 1517
QY 712 TyrThrArgThrGluGluSerAlaArgGlyLeilePheThrLysThrArgGlnSerAla 731
   : : : : :
Db 1518 -----GGAATTATTTTGTGAAAGAGAGATACACAGCA 1550
QY 732 TyrAlaLeuSerGlnTrpPheThrGluAsnGluLys-----PheAla 745
   : : : : :
Db 1551 GTTGCTTTAAACAGATTGATTAAGAGAGCTGGCAAAACAGATCCAGTACGCTTATATC 1610
QY 746 GluValGlyValLysAlaHisLysLeu1LeGlyAlaGlyHis----- 759
   : : : : :
Db 1611 AGTAGCAATTTTCATACCTGACATGCGATGGAGAGATCAGCTTCGCAATACACAGATG 1670
QY 760 SerSerGluPheLysProMetThrGlnAsnGluGluLysGluVal1LysLysPheArg 779
   : : : : :
Db 1671 GAAGCAGAAATTCAG-----AAACAGAGAGAGTACTTAAATAATTCGA 1715
QY 780 ThrGlyLys1LeaSnLeuLeu1LeaLThrValAlaGluGluLysLysAsp1LeLys 799
   : : : : :
Db 1716 GCACATGACAGACCACTGCTTATGCAACAGATTTCTACAGAGCTGTGTATATACCA 1775
QY 800 GluCysAsn1LeaVal1LeaArgTyrGlyLeuVal1ThrAsnGlu1LeaLysValGlnAla 819
   : : : : :
Db 1776 AATTCACACTGTGCTGCTTGTGATTTGCCACAGAAATACATCATCTCTTCAATCT 1835
QY 820 ArgGlyArgAlaArgAlaAspGluSerThrTyrValLeuValAlaHisLysGlySerGly 839
   : : : : :
Db 1836 AAGGAGAGGAGCAAGGCCCATCTCTAATATATATATGTTACCG----- 1880
QY 840 Val1LeGluHisGluThrValAlaAsnAspPheArgGluLysMet---Met1LysAla1Le 858
   : : : : :
Db 1881 -----GATACAGACAAATAAAGTTTGAAGAGACCTTAAGCTTAAAGCTATT 1934
QY 859 HisCysValGlnAsnMetLysProGluGluTyrAlaHisLys1LeuGluLeuGlnMet 878
   : : : : :
Db 1934 ----- 1934
QY 879 GlnSer1LeMetGluLysLysMetLysThrLysArgAsn1LeAlaLysHisTyrLysAsn 898
   : : : : :
Db 1935 -----GAAAGATCTTGAGAAACACAG----- 1955
QY 899 AsnProSerLeu1LeThrPheLeuCysLysAsnCysSerValLeuAlaLysSerGlyGln 918
   : : : : :
Db 1956 -----TGTTCCAGTCGCTTGA- ACTGCTGAG 1982
QY 919 ---Asp1LeHisVal1LeGluLysMetHisHisValAlaSerThrProGluPheLysGln 937
   : : : : :
Db 1983 ACTGACATTCATCTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2027
QY 938 LeuTyr1LeValAlaArgGluAsnLysAla----- 946
   : : : : :
Db 2028 ---TATGCTTGAGCGCTGACGATGCTGCTCCAGAGTCACATCAACATCAACATCAACAT 2084
QY 947 ---LeuGlnLysLysCysAla-AspTyrGlnLe-----As 957

```

```

Db 2085 CACATCAATAGATAGTACTGCTAGATATACCAAGTATCGCTTACTGATCTAGCT-CTTAA 2144
   : : : : :
QY 957 nglyGlu-----1LeleCysLysCysG 965
   : : : : :
Db 2145 TGCAAGACCCGAGAGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2204
QY 965 yGlnAlaTrp---GlyThrMetLeuValHis-----LysG 976
   : : : : :
Db 2205 TCACCTTCGAGCCTCCATCTGTTGTCACCAATGAGCTGTGACATTTGCTTAAAG 2264
QY 976 yLeuAspLeuPro-----CysLeuLys1LeaArgAsnPheAlaVal1Val 990
   : : : : :
Db 2265 GTTGCGCTCATTTGCTGTGAGAACTGCACAAATTTGGGAACTGATGATGATGATGAT 2322

RESULT 3
US-08-143-576-6
Sequence 6, Application US/08143576
Patent No. 5643761
GENERAL INFORMATION:
APPLICANT: Fisher, Paul B.
APPLICANT: Jiang, Hongping
TITLE OF INVENTION: METHOD FOR GENERATING A SUBTRACTED
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESS: John P. White, c/o Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/143,576
FILING DATE: 25-OCT-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 43563/JPW/AKC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 664-0525
TELEX: 422523 COOP UT
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 193 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-143-576-6

Alignment Scores:
Pred. No.: 2.3e-17 Length: 193
Score: 227.50 Matches: 50
Percent Similarity: 75.36% Conserved: 2
Best Local Similarity: 72.46% Mismatches: 10
Query Match: 4.28% Indels: 8
DB: 1 Gaps: 1

US-09-515-363c-2 (1-1025) x US-08-143-576-6 (1-193)
QY 947 LeuGlnLysLysCysAlaAspTyrGln1LeaSnGluLysLysLysCysLysCysLysGln 966
   : : : : :
Db 1 CTGCAAAAGAACTGTGCGCAGCTATATA-ATAATGTGGAATCATCTGCAAAATGTGGCAG 59

```


Db 464 TTTCAGAACACTGCGACATATGTTAAAACTACAGAGACATTGCAAAATTAG 523
 Oy 123 ProthLeuValAspLysLeuLeu---ValArgAspValLeuAspLysCysMetGluGlu 141
 Db 524 CCAAGGCTTGAGAGCAAAATTTATATCTGCAGAAAGCAATTAGACCTACTACTGTATA 583
 Oy 142 GluLeuLeuThrIleGluAspArgAsnArgIleAlaAlaIleGluAsnAspLysAsnGlu 161
 Db 584 AAGAACGAAACAGATTCTCACTCCAAATATCATGTAGAGATCAATGCTCCAGATTACAA 643
 Oy 162 SerGluValArgGluLeuLeuLysArgIleValGlnLysGlnAsnTrpPheSerAlaPhe 181
 Db 644 AAAAATATTATTAAGTTTGCAGAGAGAGCTTTACAG----- 679
 Oy 182 LeuAsnValLeuArgGlnThrGlnAsnAsnGluLeuValGlnGluLeuThrGlySerAsp 201
 Db 680 TTGAAGCTATACACCAAGACAGCTGAAAGAGCTGATGCTGACGATTAAGCATCA--- 736
 Oy 202 CysSerGluSerAsnAlaGluIleGluAsnLeuSerGluValAspGlyProGluValGln 221
 Db 737 GCTAAGCAACATGAAACAGACAGATTAATTAAGTTGAACAGAGTAAACAGACATTAAGTAAA 796
 Oy 222 Glu---GlnLeuLeuSerThrThrValGlnProAsnLeuLeuLysGluValTrpGlyMet 240
 Db 797 CATGTGAGGCAAGTGAAGAACATCCAGAGAAATATCAATGTGACTTAAGAAATTA 856
 Oy 241 GluAsnAsnSerSer-----Glu 246
 Db 857 AGCAAAAGCCACCTCAATCCAAACCAAGACATCAGATATGTCTATCTTCGCAAGAA 916
 Oy 247 SerSerPheAlaAspSerSerValValSerGluSerAspThrSerLeu----- 262
 Db 917 AATACATTTGTGAA---CAAGTAGTAATGAAAAAGTCAACACCTTAGACAGTACTCTTA 973
 Oy 263 ---AlaGluLysSerValSerCysLeuAspLysSerLeuLysLysIleAsnSerAsnMet 280
 Db 974 AAGAATTGAACTCTCAACACAGATCTCTTAAAGATGAGTAGTAACCTATATACATAATCTT 1033
 Oy 281 GlySerAspSerGlyThrMetGlySerAspSerAspGluGluAsnValAlaAlaArgAla 300
 Db 1034 ---AAGTTAAACTTGAAATGATGCTCAACATATTAAGATGAGTTT 1078
 Oy 301 SerProGluProGlu---LeuGlnLeuArgProTyrGlnMet----- 313
 Db 1079 TTTCATGAACGCGAAGACTTAGAGTTTAAATTAATGAATATATCTACTAGTAAAGACAA 1138
 Oy 314 ---GluValAlaGlnProAlaLeuGluGluLysAsnIleIleLys 328
 Db 1139 CAGGCGTGTGTAATGAAAAATTAATCTGAGCTGACAGGTTTAAATAAACGTTTGC 1198
 Oy 329 LeuProThrGlySerGlyThrArgValAlaValTyrIleAlaLysAspHisLeuAsp 348
 Db 1199 ---TATACISTAGAACACATTAACATA 1222
 Oy 349 LysLysLysLysLysLysLysLysLysLysLysValIleValLeuValLysValLeu 367
 Db 1223 GAAGTACAGAGCTTAAAGACAAACATCAAAAAGAAATATACAAATTAAGACAAATTT 1282
 Oy 368 LeuValAlaGlnLeuPheArgLysGluPheGlnProPheLeuLysLysTrpTyrArgVal 387
 Db 1283 TTGTCTGATTGCAAAAAAAGAAAA-----TTAACATTAAATGTTTGAATA 1327
 Oy 388 IleGlyLeuSerGlyAspThrGln---LeuLysIleSerPheProGluValValLysSer 406
 Db 1328 CAGGCTCTTAAGCAACAGTGTGAAACCTACACCAAGAAAGCAAGCAAGATTTAAAT 1387
 Oy 407 CysAspIleIleIleSerThrAlaGlnIleLeuGluAsnSerLeuLeuAsnLeuGluAsn 426
 Db 1388 TATGAG-----AGTTTACAGAGATTTATGAA-----ATTTTACAAACAGACAGT 1432
 Oy 427 GlyGluAspAlaGlyValGlnLeuSerAspPheSerLeuIleIleIleAspGluCysHis 446
 Db 1433 GGGGAATCTGCTCGAAAAAATTAAGTCAAGAGTTT----- 1465

Oy 447 HisThrAsnLysGluAlaValTyrAsnAsnIleMetArgHisTyrLeuMetGlnLysLeu 466
 Db 1466 GATCAATATGAACACACACACACACATCTCATGTT-----CATGAACCTCAGACAGAGTCTC 1519
 Oy 467 LysAsn-----AsnArgLeuLysLysLys 474
 Db 1520 AGACCTGCTTTACTGAAAAAGATGCCCTTCTCCAAAGCTGGAATGCCCTCCAGAGCAAA 1579
 Oy 475 AsnLysProValIleProLeuProGlnIleLeuGlyLeuThrAlaSerProGluValGly 494
 Db 1580 AATGAAAGTTACTATCTCAACACCAAGATTGTA-----CCAGAAATTGAA 1624
 Oy 495 GlyAlaThrLys-----GlnAlaLysAlaGluGlnHisIleLeuLysLeuCysAlaLysn 512
 Db 1625 AATACCATTAAGAACCTTCAAGAAAAAGAAATGAGATATCTACTACTAGTCTC---AGTCAA 1681
 Oy 513 LeuAspAlaPhe-----ThrIleLysThrValLysGluAsnLeuAsp 526
 Db 1682 ACAGATACCATGTTAAAGAAATTAGACAGAAAGATTAATTTCTCTACTAGAGAAATAAT 1741
 Oy 527 GlnLeuLysAsnGlnIleGlnGluProCysLysLysPheAlaIleAlaAspAlaThrArg 546
 Db 1742 GATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1780
 Oy 547 GluAspProPheLysGluLys-----LeuLeuGlnLeuThr 559
 Db 1781 ---GATTAATTTCCATTAAGCAATATCTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1837
 Oy 560 ArgIleGlnThrTyrCysGlnMetSer----- 568
 Db 1838 AAGATAGACCAAAACATTCACATGACACAGTACAAAGTACAAAGCAAGTAAATGAATTAACA 1897
 Oy 569 ---PrometerAspPheGlyThrGlnProTyrGluGluTrp 581
 Db 1898 GGAGGACTAGACAGACTTTAAAGAAAGATCAAAATTAATGACAAATAAGCAAAATTT 1957
 Oy 582 AlaIleGlnMet----- 585
 Db 1958 ATGCTCAATATAAGCTCTCTCTCAACCAAGAAAGATTTCTGAGTGAAGTCAAGT 2017
 Oy 586 ---GluLysLysAlaAlaLysLys----- 592
 Db 2018 CTTATGAGAAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2077
 Oy 593 ---GlyAsnArgLysGluArgValCysAlaGlnHisLeuArgLysTyrHisGlu 609
 Db 2078 GTTTTCTGCTCAAAAGAAAGATGTATCTTAAAGACATATTAATTAATTAATTAATTAAT 2137
 Oy 610 AlaLeuGlnIleAsnAspThrIleArgMetIleAspAlaTyrThrHisLeuGluThrPhe 629
 Db 2138 AAATCTTCACTTA-----ATGTTGAAGCAAGATTAATTAATTAATTAATTAATTAAT 2182
 Oy 630 TyrAsnGluLysLysLysLysPheAlaVal----- 640
 Db 2183 CTGAAATAAGACAGATTCAGAAATTTATTTAAACAGTATGATGTTTCTTAA 2242
 Oy 641 ---IleGluAspAspSerAspGluGly----- 649
 Db 2243 GAAATGCGATCAGAAAGTTTCAAGACAGTGAAGAAAGATGTTTAATGCTCAAG 2302
 Oy 650 ---AspAspGluTyrCysAsp-----Gly 656
 Db 2303 GCACTGGGTGAATCCTTGCAAAAAATTAATGAGAAAAATGCAACTGCTTTTCAAGCT 2362
 Oy 657 AspGluAspGluAspAspLysLysProLeuLysLeuAspGluThrAspArgPhe--- 675
 Db 2363 GATGAAAAAGTATGAGTTAGAAAAAGATTAAGTGCCTTCAAGCAAGAGTATCT 2422
 Oy 676 ---LeuMetThrLeuPhe-----PheGluAsnAsnLysMetLeuLysArg 689
 Db 2423 CAGTGTGAAGAACTTAAGCTTTATTGAGAGACTATGAGCAAGAAAGATTCCTTAAG 2482

QY 690 LeuAlaGluAsnProGluTyrGluAsnGluLys-----LeuThrLys 703
 DB 2483 ---AAAGACTGACAGAAATATACGCTGAGAAAGAGCCCTGAGCTGTAATCTTACAGAA 2539
 QY 704 LeuArgAsnThrIleMetGluGlnTyrThrArgThrGluGlnIleSerAlaIleGlyIleIle 723
 DB 2540 ATGACAGATGCT-----AATGAAAAACAGCCTTGAATATAGAAAT-----CTT 2584
 QY 724 PheThrLysThrArgIleSerAlaTyrAlaLeuSerGlnTrpIleThrGluAsnGluLys 743
 DB 2585 TTATTCAGAGTTGAAAGATATCTCAACATGACGAAAAAGTGAATTCATATGAAAAA 2644
 QY 744 PheAlaGluValGlyValIleLysAlaIleHisIleIleGlyAlaGlyIleSerGlyPhe 763
 DB 2645 ---GAAAAATGTTTATATAAGAACAT-----GAAAACTTA 2677
 QY 764 LysProMetThrGlnAsnGluGln-----Lys 772
 DB 2678 AACCCACTACTACACAAAAAGAAATTACGAGATAGAGACAGACAGTTGAACTATTAAAG 2737
 QY 773 GluValIleSerLysPheArgThrGlyLysIleAsnLeuIleAlaIleThrValAla 792
 DB 2738 GATTCCTTACAAAAATCACCCTCTGTAAAAAATGATCTCTCTCTCAGTAAAGAGCTG 2797
 QY 793 GluGluGlyLeuAsp-----IleLysGluCysAsnIleValIleArgTyrGlyLeuVal 810
 DB 2798 GAAGAAAAATACAAAAATCTCGAAAAAGATGCAAA-----GAAAGACAGAGAA 2848
 QY 811 ThrAsnGluIleAlaMetValGlnAlaArgGlyArgAlaArgAlaAspIleSerThrTyr 830
 DB 2849 ATAAATTAAGATTAATTAATAGTGTCCGTAAAGGCAAAAGAACTAATTTAGCAAGAAA 2908
 QY 831 ValLeuValAlaHisSerGlySerGlyValIleGluHisGluThrValAsnAspPheArg 850
 DB 2909 -----GAGACCCAGACCTGTGAAGCAAGAACTT 2935
 QY 851 GluLysMetMetTyrLysAlaIleHisCysValGlnAsnMetLys----- 865
 DB 2936 GAATCTCTTGATCAGAAAAAGACCCAGTATCTGCTCCATGAGACATCTATTACAGA 2995
 QY 866 ProGluGluTyrAlaHisLysIleLeuGluLeuGlnMetGlnSerIleMetGluLysLys 885
 DB 2996 GCAGAAAGCTATTAAGATCTTTATTAGAAATATGAAAGACTCA-----GAGCAA 3046
 QY 886 MetLysThrLysArgAsnIleAlaLysHisTyrLysAsnAsnProSerIleIleThrPhe 905
 DB 3047 CTGGATGTGGAAAAAGAACCTGCTAATAATTGAGCATGTATTGAAAGACTTACAGCA 3106
 QY 906 LeuCysLysAsn-----CysSerValLeuAlaCysSerGlyGluAspIleHis 921
 DB 3107 CAATTAGAAATTCGACTTTCGAGTGTGAACCAATTAATCTATATGACGATCTC--- 3163
 QY 922 ValIleGluLysMetHisHisValAsnMetThrProGluPheLysGluTyrIleVal 941
 DB 3164 ---CTGGCTGCTGATATGACATATACAGTCTAATGCCAAATTATTAAGATGATTTTA 3220
 QY 942 ArgGluAsnLysAlaLeuGlnLysCysAlaAspTyrGlnIleAsnGluIleIle 961
 DB 3221 GAAGTCACAGAGCC-----AAAGCAATGCTGACCAAGAAATTAAGCTGAAAACTT 3274
 QY 962 CysLysCysGlyGlnAlaIleTyrGlyThrMetMetValHisLysGlyLeuAspLeuPro 981
 DB 3275 CAGAAAGAACAGAAATA-----AAGAACATCTCTACTATCTGTA 3313
 QY 981 IleuLysIleArgAsnPheValValPhe-----LysAsnAsnSerThrLysLysGlu 999
 DB 3314 AATGAACCTTCAACACTTCTGCTACAACTTCAAAAAAGCAACAGACCTTCAGAAAAAC 3373
 QY 999 nTyrLys-----LysTyrValGluLeuProIleThrPheProAsn 1012
 DB 3374 ATGCAAGAAATTAAGCTGGTTAA-----AGATGCTGCAAC 3410
 RESULT 6

US-08-139-937-12
 / Sequence 12, Application US/08139937
 / Patent No. 5821070
 / GENERAL INFORMATION:
 / APPLICANT: LEE, MEN-HWA
 / APPLICANT: SHAN, BEN
 / TITLE OF INVENTION: CELLULAR GENES ENCODING
 / TITLE OF INVENTION: RETINOBLASTOMA-ASSOCIATED PROTEINS
 / NUMBER OF SEQUENCES: 14
 / CORRESPONDENCE ADDRESS:
 / ADDRESSEE: CAMPBELL AND FLORES
 / STREET: 4370 LA JOLLA VILLAGE DRIVE
 / CITY: SAN DIEGO
 / STATE: CALIFORNIA
 / COUNTRY: USA
 / ZIP: 92122
 / COMPUTER READABLE FORM:
 / MEDIUM TYPE: Floppy disk
 / COMPUTER: IBM PC compatible
 / OPERATING SYSTEM: PC-DOS/MS-DOS
 / SOFTWARE: Patent In Release #1.0, Version #1.25
 / CURRENT APPLICATION DATA:
 / APPLICATION NUMBER: US/08/139,937
 / FILING DATE: 20-OCT-1993
 / CLASSIFICATION: 435
 / PRIOR APPLICATION DATA:
 / APPLICATION NUMBER: US 07/979,156
 / FILING DATE: 20-NOV-1992
 / ATTORNEY/AGENT INFORMATION:
 / NAME: CAMPBELL, CATRYN
 / REGISTRATION NUMBER: 31,815
 / REFERENCE/DOCKET NUMBER: P-CJ 9370
 / TELECOMMUNICATION INFORMATION:
 / TELEPHONE: 619-535-9949
 / TELEFAX: 619-535-9001
 / INFORMATION FOR SEQ ID NO: 12:
 / SEQUENCE CHARACTERISTICS:
 / LENGTH: 4868 base pairs
 / TYPE: nucleic acid
 / STRANDEDNESS: single
 / TOPOLOGY: linear
 / MOLECULE TYPE: cDNA
 / US-08-139-937-12
 /
 / Alignment Scores:
 / Pred. No.: 1,14e-11 Length: 4868
 / Score: 199.50 Matches: 219
 / Percent Similarity: 35.52% Conservative: 187
 / Best Local Similarity: 19.16% Mismatches: 472
 / Query Match: 3,768 Indels: 265
 / DB: 1 Gaps: 46
 /
 / US-09-515-363C-2 (1-1025) x US-08-139-937-12 (1-4868)
 /
 QY 21 ArgValLysMetTyrIleGluValGluProValLeu----- 32
 DB 207 AAAATTGAAGCATGATGAATGGAAAAATGTTGGGCACTTAAGAAAGAAATTC 266
 QY 33 -----AspTyrLeuThrPheLeuProAlaGluValLysGluGlnIleGlnArgThr 49
 DB 267 GATTAAAGTGAATAATTTGAAATATTTCTGTGATCAACAGAGATTTACTGACAGACTA 326
 QY 50 ValAlaThrSerGly-----AsnMetGlnAlaValGluLeuLeuLeuSer 64
 DB 327 GAACCTTGAAGGCCCTCAATTCGATTAGAAATATGATGACAGATTAATCATACAGTGA 386
 QY 65 ThrLeuGluLysGlyValIleThrHisLeu-----GlyTyrThrArgGluPheValGluAla 82
 DB 387 GATATTGAGATAAATGCGCCAGCTGATGACACAGCTGAAGAGAGCATTTCTTAT 443
 QY 83 LeuArgArgThrGlySerProLeuAlaIleArgTyrMetAsnProGluLeuThrAspLeu 102
 DB 444 -----GTGAAATATGAGCTGTGAGTAGATC 467

QY 103 ProSerProSerPheGluAsnAlaHisAspGluTyrLeuGluLeuLeuAsnLeuGln 122
 DB 468 AGATGGAGACAAAGCTAGCATTTAGCAT---GAAGCCCTTACCTTGAGCTTACCTTAGC 524
 QY 123 ProThrLeuValAspLysLeu---LeuValArgAspValLeuAspLys----- 137
 DB 525 GTAGTCAACAAAGAGAGAGTATGTTAGAAAAGCAATGAAATATGAAAGAGTATT 584
 QY 138 ---CysMetGluGluGluLeuLeuThrIle---GluAspArgAsnArgIleValAla 155
 DB 585 GTCTGCTTGAAGAGAACTCTGAGTGCACAAAGTGAAGAAACAGCTTGTGTGAGAA 644
 QY 156 GluAsnAsnGlyAsnGlu---SerGlyValArgGluLeuLeuLysValLeuGln 173
 DB 645 TAGATCTACTATGCAAAAAAACCAGGACCTGATGATGTTGTCTGAAAATGAAAGAG 704
 QY 174 Lys-----GluAsnTrpPheSerAlaPheLeuAsnValLeuArgGlnThr--- 188
 DB 705 AAACACAAAGAGCTTGAAGCTCATCAAAAGTGAAGTCTCTCATTTGCATTCAGCTGCGAG 764
 QY 189 -----GlyAsnAsnGluLeuValGlnGluLeuThrGlySerAspCysSerGln 204
 DB 765 GCAGAGGTGAAGAAAAGACGGAAGCTTCTGAGACTTGTGCT---TCTGATGAGAGTGA 821
 QY 205 SerAsnAlaGluIleGluLeuAsnSerGln---ValAspGlyProGlnValGluGln 223
 DB 822 CTGTTAAAGACAAACTATCTCCAGAAAGCTGCAAGTTTGAAAGAGCTCAGCAG 881
 QY 224 LeuLeuSerThrThr-----ValGlnProAsnLeuGlnLys 235
 DB 882 GCACCTCTTGTGCAAAATGTGAGCTGGAAGAAACCAATTTGCACACTGATGAAAGGAA 941
 QY 236 GluValTyrGlyMetGluAsnAsnSer-----SerGlnSerPheAla 250
 DB 942 GAATTCCTTGTCAAGAACTGAAAGCCCTGCAAGCCAGACTGAGTAATAGATTATGAA 1001
 QY 251 AspSerSerValSerGluSerAspThrSerLeuAlaGluGlySer----- 266
 DB 1002 AACCTAAATGTCTCCAGAGCCCTTGAAGCCGCACTGTGTGGAAGAGTTCGATTCG 1061
 QY 267 ---ValSerCysLeuAspGluSerLeuGlyHisAsnSerAsnMetGlySerAspSerGly 285
 DB 1062 AGCGTAGCTCAACACAGAGGAGGAGT---CATCACCTGAGAAAGAGCATCGAGAACTG 1118
 QY 286 ThrMetGlySerAspSerAspGlu-----GluAsnValAlaAla 298
 DB 1119 AGAGTTCGATTCGAGCCGCTGAAAGAAAGCACTGACATGCAAGAGAACTGAAAGAA 1178
 QY 299 ArgAlaSerProGluProGluLeuGlnLeuArgProTyrGlnMetGlnValAlaGlnPro 318
 DB 1179 CGCGACGGGAGACATATTCATCTTAAGGTAAAGTTGACAACTTCAAAATGCAATTCGAC 1238
 QY 319 AlaLeuGluGlyLysAsnIleLeuLeuLeuLeuProThrGlySerGlyValAla 338
 DB 1239 ATGTCAAGAAACAAAGAGAGCTAGTCTTGTGATGCGAGATTCGAAAGAGAGAGTA 1298
 QY 339 AlaValTyrIleAlaLys---AspHisLeuAspLysLysLysLysAlaSerGluProGly 357
 DB 1299 GAGACTCTAAAAACCAAAATAGAGAGATGGCCGAGAAAGTGAAGTTTGTGATTAAGAC 1358
 QY 358 LysValIleValLeuValAsnLysValLeuLeuValGlnGluLeuPheAsnGlyGluPhe 377
 DB 1359 CTTCGACGTTAAGTCTGAAAGAAATCTGACAAACAAATACAAAGAAACAGGT 1418
 QY 378 GlnProPheLeuLysLysTyrArgValIleGly-----LeuSerGly 392
 DB 1419 CAG-----TTGTCAAGACAGTACAGTACTCTTCAATTTAAGTCTGTTAGAGAA 1472
 QY 393 AspThrGlnLeuLysIleSerPheProGluValValLysSerCysAspIleIleLeuSer 412
 DB 1473 AAGGACAGGAGAGATACAGATCAAAAGAAATATCAAA-----ACT 1514

QY 413 ThrAlaGlnIleLeuGluAsnSerLeuLeuAsnLeuGluAsn----- 426
 DB 1515 GCAGTGGAGATGCTTCAAGATCAGTAAAGAGAGCTAAATGAGGAGTACAGCCCTTCGT 1574
 QY 427 GlyGluAspAlaGlyValAlaGlnLeuSerAspPheSerLeuIleIleIleAspIleLysHis 446
 DB 1575 GGTGACCAAGAAATATATGACAGCCACAGACAGACTCTAGACCCCAATTAAGAGAGAG 1634
 QY 447 HisThrAsnLysGluAlaValTyrAsnAsnIleMetArgHisTyrLeuMetGlnLysLeu 466
 DB 1635 CATCAGCTGCAAAATAGATT-----GAAAGAGCTG 1664
 QY 467 LysAsnAsnArgLeuLysLysGluAsnLysProValIleProLeuProGlnIle----- 484
 DB 1665 AGAGCCCGCTTAGAGCTGATGATAAAGAGACAGCTGTGTCTTCAACAACTGAGGAA 1724
 QY 484 ----- 484
 DB 1725 AGTAGCATCATGAGATTACTTAAGGTTAGTGAAGCAACTGAAAGAGCTAGAG 1784
 QY 485 LeuGlyLeuThrAlaSerProGlyValGlyGlyAlaThrLysGlnAlaLysAlaIleGlu 504
 DB 1785 ATAGCCAGACAAACCAAGAGCATGACCTCTTGAAGGAGAGAAATCCAAAGAGCTA 1844
 QY 505 HisIleLeuLysLeu-----CysAlaAsnLeuAsp 514
 DB 1845 GAGACCTTAAGGCAAAATAGAGAGGATGACCCAAAGCTGAGAGCTCTGAAATACAT 1904
 QY 515 AlaPheThrIleLysThrValLysGluAsnLeuAspGlnLeuLysAsnGlnIleGlnGln 534
 DB 1905 GTTGTACTATAGGTTCAGAAAAGAAATATGCAAAATGCAATTCACAAAGAGAG 1964
 QY 535 ProCysLysLysPheAlaIleAlaAspAlaThrArgGluAspProPheLysLysLeu 554
 DB 1965 CGAATATCTGAATAGAAATTAATTAATTCATATTGAAATATTTCCAAAGAA--- 2021
 QY 555 LeuGluIleMetThrArgIleThrTyrCysGlnMetSerProMetSerAspPheGly 574
 DB 2022 -----GAGCAAGCAAAAGTACAGATGAAAGAAATCAAGACCTGCTC 2063
 QY 575 ThrGlnProTyrGluGlnThrAlaIleGlnMetGluLysLysAlaLysLysLysAsn 594
 DB 2064 ATGAGAGATGCTTCAACAAACAAATTAAGAGGCTCATGAGAGAGTGCAGCCCTGATAT 2123
 QY 595 ArgLysGluArgValCysAla-----GluHisLeuArgLysTyrAsnGluAlaGln 612
 DB 2124 GACCAAGAGCCCTGTAGGCCCAAGAGCAAGTCTTACTAGTCAAGTAAAGTGTGAA 2183
 QY 613 IleAsn-----AspThrIleArgMetIleAspAla-----TyrThrHisLeu 626
 DB 2184 CTTCAGAAAGCTCAGTCTGCTACAGAGCCCTTGTAGAGCCCAAAATATATATATTTG 2243
 QY 627 GluThrPheTyrAsn-----GlnGln 633
 DB 2244 CATCTTCAGTGAATGGCTCATTCAGAGATGAGAAGATGGCAAGCAAGAACTGAGAGAG 2303
 QY 634 LysAspLysLysPheAlaValIleGluAspAspSerAspGluGlyGlyAspAspThrLys 653
 DB 2304 AAGGATGAAGAAATAGTACAGTGAATAATCAATT----- 2339
 QY 654 CysAspGlyAspGluAspLysAspLysLysProLeuLysLeuAspGluThrAsp 673
 DB 2340 -----CAAGACCAAGAGAGCTTGTCTTAAACTGTCCAGAGTGAAGAGAG 2387
 QY 674 ArgPheLeuMetThrLeuPhePheGluAsnAsnLysMetLeuLysArgLeuIleAlaAsn 693
 DB 2388 CAC-----CACTTTCGAAGAGAGCAAACTTAGAAGCTGAGAAATCTGACAGAGAGAA 2438
 QY 694 ProGluTyr-----GluAsnGluLysLeuThrLysLysLeuAsn 706
 DB 2439 TTGAGCAGAGATCAACAGTCTACAAATCCAAATGCTCTTTCGAGAGACATTAAGAA 2498
 QY 707 ThrIleMetGluGlnTyrThrArgThrGlnLysSerAlaArgGlyIleIlePheThrLys 726


```

OY 767 ThGlnAsnGluGlnIleGlyValIleSerLysPheArgThrGlyLysIleAsnLeuLeu 786
DB 2646 GAGCTGCAAGAAAGAACTGAGTGGAGAGAAAAATAGGCTAGCTGAGAGAGTGGAGTACTG 2705
OY 787 -----IleValThrThrVal 791
DB 2706 TTGGAAAGAAATTAAGAGCAGCAAGATCAATTAAGAGAGCTTACACCTAGAAATTAAGTCAA 2765
OY 792 AlaGluGluGlyLeuAspIleLysGlyCysAsnIleValIleArgThrGlyLeuValThr 811
DB 2766 TTGAGAGAGAGCTAGATTTCATGCAAGAAAGAGAGTGGAAAGAGAGAGAGAGAGAGAGAG 2825
OY 812 AsnGluIleAlaMetValGlnAlaArgGlyArgAlaArgAlaAspGluThrThrVal 831
DB 2826 GAGGAAATAGCTGATATTCAGCTACGCTTCATGAGAGCTGAAAGAGAGAGAGAGAGAGCTTGG 2885
OY 832 LeuValAlaIleSerGlySerGlyValIleGluHisGluThrValAsnAspPheArgGlu 851
DB 2886 CTTTGGACACAAACAAACAGTGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2936
OY 852 LysMetMetLysLysAlaIleHisCysValGlnAsnMetLysProGlu----- 867
DB 2937 AATTGACTCTTAAGAA-----GAATGCTGAGTTCACAGAGAGAGAGAGAGAGAGAGAGAG 2993
OY 868 -----GluThrAlaHisLysIleLeuGluLeuGluMetGluSerIleMetGlu 883
DB 2994 AAGTCTAGTAAGAGAGAGCTGATTAATTCATGAAAGTACTACTGCA-----ATTGGGAA 3050
OY 884 LysLysMetLysThrLys-----ArgAsnIleValLysHisLysIle 896
DB 3051 GAATTGACAAACAAACAGATGAGCAATCTAATAATGTAATGTAATGTAATGTAATGTAATGTAAT 3110
OY 897 LysAsnAsnProSerLeuIleThrPheLeuGlyLysAsnGlySerValLeuAlaCysSer 916
DB 3111 GAACGTGCCAGGAGGAGAAATGAAAGTGTGATCAAAATCTGTAAACAGAGCTGGAA----- 3164
OY 917 GlyGluAspIleHisValIleGluLys----- 925
DB 3165 ---GAGGAAAGAGATGACTGAGAGAAAGAACTCTCAAGCTGCAAGCTGCAAGAGAGAG 3221
OY 926 -----MetHisIleValAsnMetThrProGluThrPheLysGlu 937
DB 3222 CAGAAAGAGAGTACTGTTATGATACCAAGAGTGCATGATTAACAACTGATCAAGAGAGAG 3281
OY 938 LeuThrIleValArgGluAsnLysAla-----LeuGluLysLysCys 951
DB 3282 CTGAAAGAAAGCTCTGAAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3341
OY 952 AlaAspThrGlnIleAsnGlyGluIleIleGlyLysCysGlyGlnAlaIleArgThrMet 971
DB 3342 TCC---TTGCTTAAAGCATGAAAGATTAGAGAAAGCTAAACAGATGTAAGAGAGAGAG 3398
OY 972 MetValHis 974
DB 3399 GTGCCCAT 3407

```

```

RESULT 8
US-08-328-254-5
; Sequence 5, Application US/08328254
; Patent No. 5710022

```

```

GENERAL INFORMATION:
APPLICANT: Zhu, Xueliang
APPLICANT: Lee, Wen-Hwa
TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/328-254
FILING DATE: 24-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/141,239
FILING DATE: 22-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-CJ 1191
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 8789 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 544..7990
US-08-328-254-5
Alignment Scores:
Pred. No.: 1.22e-10 Length: 8789
Score: 194.50 Matches: 219
Percent Similarity: 35.43% Conservative: 186
Best Local Similarity: 19.16% Mismatches: 473
Query Match: 3.66% Indels: 265
DB: 1 Gaps: 46
US-09-515-363c-2 (1-1025) x US-08-328-254-5 (1-8789)
OY 21 ArgValLysMetThrIleGlnValGluProValIle----- 32
DB 4108 AAAATTGAAAGCATGATGAAATTTGGGAAAGTGAAGAAAGAAAGTCA 4167
OY 33 -----AspThrLeuThrPheLeuProAlaGluValLysGluGlnIleGluArgThr 49
DB 4168 GATTAAAGGAAATTTGAAATATTTCTGTGATCAAGAGATTAATCTCAGAGAGTCA 4227
OY 50 ValAlaThrSerGly-----AsnMetGlnAlaValGluLeuLeuSer 64
DB 4228 GAACTTGAAGAGGCTCAATTCGATTAGCAAAATGCAATGCAATGCAATGCAATGCAATGCAAT 4287
OY 65 ThrLeuGluGlyValIleThrPheLys-----GlyThrThrArgGluPheValGluAla 82
DB 4288 GATATTGGAGATAATGTCGCCAAGGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4344
OY 83 LeuArgArgThrGlySerProLeuAlaIleArgGlyPheAsnProGluLeuThrAspLeu 102
DB 4345 -----GTGCAAAATAGAGCTGAGTAGCATC 4368
OY 103 ProSerProSerPheGluAsnAlaHisAspGluThrLeuLeuLeuAsnLeuLeuGln 122
DB 4369 AGATGCAAGAAAGTACGATTCAGCAT---GAAAGCCTCTACCTGAGAGCTGATACGATGAG 4425
OY 123 ProThrLeuValAspLysLeu-----LeuValArgAspValLeuAspLys----- 137
DB 4426 GTAGTCAACAGAGAACTGATTGTAAGAAACCAATGAAATTAACAGAGAGAGTATT 4485
OY 138 ---CysMetGluGluGluLeuLeuThrIle---GluAspArgAsnArgIleAlaIleAla 155
DB 4486 GTCTGCTTGAAGAAAGAACTCTCAAGTGTCAAGTGTCAAGTGTCAAGTGTGTGAGTAA 4545

```


[illegible]

RESULT 9
 US-08-353-700-2
 : Sequence 2, Application US/08353700
 : Patent No. 5599919
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: YEN, TIMOTHY J.
 : APPLICANT: RATTNER, JEROME B.
 : TITLE OF INVENTION: NUCLEIC ACID ENCODING A
 : TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETICPHONE PROTEIN
 : TITLE OF INVENTION: AND METHODS OF USE
 :
 : NUMBER OF SEQUENCES: 4
 :
 : CORRESPONDENCE ADDRESS:
 :
 : ADDRESSEE: DANN, DOFFMAN, HERRELL AND SKILLMAN
 : STREET: 1601 MARKET STREET, SUITE 720
 : CITY: PHILADELPHIA
 : STATE: PA
 : COUNTRY: USA
 : ZIP: 19103-2307
 :
 : COMPUTER READABLE FORM:

```

1  MEDIM TYPE: floppy disk
2  COMPUTER: IBM PC compatible
3  OPERATING SYSTEM: PC-DOS/MS-DOS
4  SOFTWARE: PatentIn file #1.0, Version #1.25
5
6  CURRENT APPLICATION DATA:
7  APPLICATION NUMBER: US/08/353,700
8  FILING DATE: 09-DEC-1994
9  CLASSIFICATION: 435
10 ATTORNEY/AGENT INFORMATION:
11 NAME: REED, JANET E.
12 REGISTRATION NUMBER: 36,252
13 TELECOMMUNICATION INFORMATION:
14 TELEPHONE: (215) 563-4100
15 TELEFAX: (215) 563-4044
16 INFORMATION FOR SEQ ID NO: 2:
17 SEQUENCE CHARACTERISTICS:
18 LENGTH: 1016 base pairs
19 TYPE: nucleic acid
20 STRANDEDNESS: double
21 TOPOLOGY: linear
22 MOLECULE TYPE: cDNA to mRNA
23 HYPOTHETICAL: NO
24 ANTI-SENSE: NO
25 ORIGINAL SOURCE:
26 ORGANISM: HUMAN
27
28 US-08-353-700-2

```

Alignment Scores:		Length:	
Pred. No.:	3 47e-10	Matches:	10136
Score:	191.50	Mismatches:	218
Percent Similarity:	35.60%	Conservative:	185
Best Local Similarity:	19.26%	Mismatches:	486
Query Match:	3.61%	Indels:	243
DB:	1	Gaps:	45

QY	21	ArgvallylsMetYrllleGlnValGluProValLeu-----	32
Db	5919	AAATTCAGACATGCATGAAATTGGAAAAAAATAGTTGGGGAACTTAACAAACAACACCA	5978
QY	33	-----AspYrllLeuThrPheLeuProAlaGluValLysGluGlnIleGlnGthr	49
Db	5979	GATTTAAGTAAAAAATTTGGAAATATTTTCTGTGATGACACAGACATTAAGTACACAACTA	6038
QY	50	ValAlaThrSerGly-----AsnMetGlnAlaValGluLeuLeuProSer	64
Db	6039	GAACCTTCGAGGCCCTCATCTGTATTAGAAATGCATGCAGATAAATCATCACTGTA	6098
QY	65	ThrLeuGluLysGlyValTrrPheLseu-----GlyTrrPthrArgGluPheValIleAla	82
Db	6099	GATTATGGACATTAATGGGGCAAGGGAATGCACAGCTGGAGAGAGATTTCTTCACT	6155
QY	83	LeuArGArGThrGlySerProLeuAlaAlaArgTyrMetAsnProGluLeuThrAspLeu	102
Db	6156	-----GTGGAAAAAGAGCTGATGATACATC	6179
QY	103	ProSerProSerPheGluAsnAlaHisAspGluTyrLeuGlnLeuLeuAsnLeuGln	122
Db	6180	AGATCGGAGAAAGCTAGCATTTGAGCAT-----GAAGCCCTTACCTGGAGGCTGATTAGAG	6236
QY	123	ProThrLeuValAspLysLeu-----LeuValArgAspValLeuAspLys-----	137
Db	6237	GTAATTAACACAGAAAGCATATTTTAGAAAAAACACAAATGAAATATACAGAAATATAT	6296
QY	138	-----CysMetGluGluLeuLeuLeuThrIle-----GluAspArgAsnArgIleAlaValAla	155
Db	6297	GTCGCTTGAAGAAACAACTCTCAGGGGTCCACAAAGAGAGAGAAACACCTTCTTCAGAA	6356
QY	156	GluAsnAsnGluLysGlu-----SerGlyValArgGluLeuLeuLysArgGluValGln	173
Db	6357	TTGATGCTATGTCAAAAAAACACAGGACATGATCACTGATCTTGAAAAAATGACAGAG	6416

174 Lys-----GluasntrpPheSerAlaPheLeuAsnValLeuArgIntr----- 188
 6417 AAACACAGAGCTGAGCTCATCAAGAGTGTCTCTGATTCAGTGTGAGCAG 6476
 189 -----GlysnAsnGluLeuValGlnGluLeuThrGlySerPheCysSerGlu 204
 6477 GCAGAGGTGAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6533
 205 SerAsnAlaGluLeuLeuAsnLeuSerGln---ValAspGlyProGlnGluGln 223
 6534 CTGTTAAAGACAAAAGTCTATCTCCAGAGAGAGAGAGAGAGAGAGAGAG 6593
 224 LeuLeuSerThrThr-----ValGlnProAsnLeuGluGln 235
 6594 GCAGTGTCTTACAAAGATGAGTGAAGAACCAATTCACAGCTGAAATAGAG 6653
 236 GluValTyrGlyMetGluAsnAsnSer-----SerGlnPhePheAla 250
 6654 GAATTCCTGTCAGAGATCTGAAGAGCTGACAGGAGAGAGAGAGAGAGAG 6713
 251 AspSerValValSerGluSerAspThrSerLeuAlaGluGlySer----- 266
 6714 AACCTGATGTCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6773
 267 ---ValSerCysLeuAspGluSerLeuGluHisAsnSerAsnMetGlyAspSerGly 285
 6774 AGGCTGAGCTCAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6830
 286 ThrMetGlySerAspSerAspGlu-----GluAsnValAlaAla 298
 6831 AGAGTTCGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6890
 299 ArgAlaSerProGluProGluLeuGlnLeuAlaArgProGluMetGluValAlaGlnPro 318
 6891 CGGAG 6950
 319 AlaLeuGlnGlyLysAsnIleIleCysLeuProThrGlySerGlyLysThrArgVal 338
 6951 ATGTCAAG 7010
 339 AlaValTyrIleAlaLys---AspHisLeuAspLysLysLysAlaSerGluProGly 357
 7011 GAGAGCTTAAAGACAAATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7070
 358 LysValIleValLeuValAsnLysValLeuLeuValGlnLeuPheAlaLysGluPhe 377
 7071 CTGTGACGTTAAGGTCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7130
 378 GlnProPheLeuLysLysTyrTyrArgValIleGly-----LeuSerGly 392
 7131 CAG-----TTGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7184
 393 AspThrGlnLeuLysIleSerPheProGluValValLysSerCysAspIleIleIleSer 412
 7185 AAGGAG 7226
 413 ThrAlaGlnIleLeuGlnAsnSerLeuLeuAsnLeuGlnAsn----- 426
 7227 GCAGTGAGAGTGTCTCAGAGATCAGTTAAGAGAGAGAGAGAGAGAGAGAG 7286
 427 GlyLysAspAlaGlyValGlnLeuSerAspPheSerLeuIleIleLeuSerGlyCysHis 446
 7287 GGTGAG 7346
 447 HisThrAsnLysGluAlaValTyrAsnAsnIleMetArgHisTyrLeuMetGlnLysLeu 466
 7347 CATCAGCTGAG 7376
 467 LysAsnAsnArgLeuLysGluAsnLysProValIleProLeuProGluIle----- 484
 7377 AGAGCCCGCTCAG 7436
 484 ----- 484

7437 AGTGAATCATGACAGATTACTTAAGGTAAGTGCAGAGAGAGAGAGAGAG 7496
 485 LeuGlyLeuThrAlaSerProGlyValGlyGlyAlaThrLysGlnAlaLys 504
 7497 ATAGCCAG 7556
 505 HisIleLeuLysLeu-----CysAlaAsnLeuAsp 514
 7557 GAGAGCTTAAAG 7616
 515 AlaPheThrIleLysThrValLysGluAsnLeuAspGlnLeuLysAsnGlnIleGln 534
 7617 GTTGTACTATAG 7676
 535 ProGlyLysLysPheAlaIleLeuAspAlaThrArgGluAspProPheLysGluLysLeu 554
 7677 CGAATATCTGAAATTAAGAAATTAATTCATGATTTGAAATATTTTGCAGAGAGAG 7733
 555 LeuGluIleMetThrArgIleGlnThrTyrCysGlnMetSerProMetSerPheGly 574
 7734 -----GAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7775
 575 ThrGlnProTyrGluGlnIleThrAlaIleGlnMetGluLysLysAlaLysLysGluLys 594
 7776 ATGAGAGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 7835
 595 ArgLysGluArgValCysAla-----GluHisLeuArgLysTyrAsnGluAlaLeuGln 612
 7836 GAGCAG 7895
 613 IleAsnAspThrIleArgMetIleAspAlaTyrThrHisLeuGlnThrPheTyrAsnGlu 632
 7896 CTT---GAG 7952
 633 GluLysAspLysLysPheAlaValIleGluAspAspSerAsp-----Glu 647
 7953 TTGCAATCTTCAAG 8012
 648 GlyLysAspAspGluTyrCys-----AspGlyAspGluAspGluAspPheLys 664
 8013 AAG 8072
 665 LysProLeuLysLeuAspGluThrAspArgPheLeuMetThrLeuPheGlnAsn 684
 8073 TCTAAAGCTGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8123
 685 LysMetLeuLysArgLeuAlaGluAsnProGluTyr-----Glu 697
 8124 TTAG 8183
 698 AsnGlnLysLeuThrLysLeuArgAsnThrIleMetGlnGlnTyrThrArgThrGluGln 717
 8184 AATGAGCTTGTGAG 8243
 718 SerAlaArgLysIlePheThrLysThrArgGlnSerAlaTyrAlaLeuSerIntrP 737
 8244 GAG-----CTTAATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8282
 738 IleThrGluAsnGluLysPheAlaGluValGlyValLysAlaHisHisLeuIleValAla 757
 8283 -----GAG 8330
 758 GlyHisSerSerGluPheLysProMetThrGlnAsnGlnLysGluValIleThrLys 777
 8331 ATGCATGAGATGAG 8390
 778 PheArgThrGlyLysIleAsnLeuLeu----- 786
 8391 AGGCTAGCTGAG 8450
 787 -----IleAlaThrThrValAlaGlnGluGlyLeuAspIleLysGlyLysAsn 802

DB 8451 AAGAGCTGACACTAGAAAATAGTGAATGAAAGAGGCTAGATTGCAATGCAAAACAG 8510
 OY 803 l1eall1leatrgtlycylleuvalthraangluilealmetvalgluinalarglyarg 822
 DB 8511 CAGCTGGAAAGGAGGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8570
 OY 823 Alatargalaspcluserthrtlyrvalleuvalalhisserglyserllyvalilcglu 842
 DB 8571 GAAGCTGAAAGAAACACAGAGGCTGTTGGACACAAACAAACAGAGAAAGAGAGAA 8630
 OY 843 Hlsgltthvalasnaphearaglyulysmetmettyllysallalhisryvalcglu 862
 DB 8631 ATCCACACA-----TACCAGAGAAATGACTTCTTAAGAA-----GAATGCTCTACT 8678
 OY 863 Asnmetlysproglu-----glutylalalhislylileu 874
 DB 8679 TCACAGACAGCTGAGATACCTTTAAAGTCTAGTAAGAGAGCTCAATATCATG 8738
 OY 875 Gluleuglimerglinserlilemetglyllysmetlysthrlys----- 889
 DB 8739 AAGGCTACTACTAG---ATTGGAAAGATTGAAAGAAACCAAGATGAGCAATCTAATA 8795
 OY 890 -----ArgAsnlllealalysHlstrlyrAsnAsnProserleuilethPheleucys 907
 DB 8796 TATGTAATGACTGAAAGAAAGAAATGAACGTCGCCAGGAGAAATGAATGAAATGTTGATC 8855
 OY 908 LysAsnGysServallleualacyserglygluaspilhlisvalileglulys----- 925
 DB 8856 AATCTGTAACAGCTGAA-----GAGGAAAGAGAGATACTGCAAGAGAAATC 8906
 OY 926 -----MetHlsls 928
 DB 8907 TCTCAACTTCAGCTGCAAGAGAGAAAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8966
 OY 929 ValasmetthProglupheyllysgluileuyllevalarqluasndysala----- 946
 DB 8967 GATGAATTAACAGCTGAGATCAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 9026
 OY 947 -----LeuGlulysLysCysAlaAspTYrGluileasnglylileilcyls 962
 DB 9027 GCAGATGAATACTTGATGATAGTACTGCTCC---TTGCTTATTAAGCCATGAAGAGTTAGAG 9083
 OY 963 LyscysgllygluinalatrrpqllythrmctmetvalHls 974
 DB 9084 AAGCTAAAGAGATGTTAGACACACAGTGGCCCAT 9119
 RESULT 10
 PCT-US95-16216-2
 ; Sequence 2, Application PC/TUS9516216
 ; GENERAL INFORMATION:
 ; APPLICANT: Yen, Timothy J.
 ; APPLICANT: Ralner, Jerome B.
 ; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
 ; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
 ; STREET: 1601 Market Street Suite 720
 ; CITY: Philadelphia
 ; STATE: PA
 ; COUNTRY: USA
 ; ZIP: 19103-2307
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US95/16216
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/353,700

FILING DATE: 09-DEC-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Reed, Janet E.
 REGISTRATION NUMBER: 36,252
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (215) 563-4100
 TELEFAX: (215) 563-4044
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 10136 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHEICAL: NO
 ANTI-SENSE: NO
 PCT-US95-16216-2
 Alignment Scores:
 Pred. No.: 3,47e-10 Length: 10136
 Score: 191.50 Matches: 218
 Percent Similarity: 35.60% Conservative: 185
 Best Local Similarity: 19.26% Mismatches: 486
 Query Match: 3.61% Indels: 243
 Gaps: 45
 US-09-515-363c-2 (1-1025) x PCT-US95-16216-2 (1-10136)
 OY 21 Argvallysmetlyrllglvalgluprovalleu----- 32
 DB 5919 AAAATTGAAGCATGATGAAATTTGAAATTTGAGGAGAACTTAAGAAATGAAATGCA 5978
 OY 33 -----AspTYrleuThrPheleuProAlagluvallysgluilileth 49
 DB 5979 GATTTAAGTGAAGAAATTTGAAATTTCTTGTGATCAGAGAGATTCTCAAGAGAGAG 6038
 OY 50 ValAlatHnsergly-----Asnmetglinalvalgluleu----- 64
 DB 6039 GAACCTTCTGAAGGCTCAATTTGATTTAGAAATGATGATGATTAATTAATTAATGCA 6098
 OY 65 ThrleuglulysgllyvaltrpHlslsleu-----GlyTrpThrArgGluPheValAla 82
 DB 6099 GATATTGAGATTAATGTGGCCAGGTGAATGACAGCTGGAAGAGAGATTTCTTAT--- 6155
 OY 83 leuArqarqthrglySerProleuAlaAlaArgTrMetAsnProgluleuThAsp 102
 DB 6156 -----GTGAAATATGACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6179
 OY 103 ProserProserPhegluasnalalhisaspilurlyrleuGluileuAsnleuileu 122
 DB 6180 ACATCGAGAGAAAGCTAGATTCAGCAT---GAGCCCTCTACTGAGAGCTGATTAAGAG 6236
 OY 123 ProThrleuValAspLysleu---LeuValArgaspValleuAspLys----- 137
 DB 6237 GTAGTCAACAGAGAGAGATGTTTGAAGAAAGCAATGAAATTAAGCAGAGATTTAT 6296
 OY 138 ---CysMetgluglululeuThrlle---GluAspArgAsnArgAlaAlaAla 155
 DB 6297 GTCTGCTTGAAGAGAGAGCTGCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6356
 OY 156 Gluasnasnglyasnglu-----SerGlyValArgGluileuLysArgAlaValcglu 173
 DB 6357 TTAGATCTATGTAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6416
 OY 174 Lys-----GluAsnTrpPheSerAlaPheleuAsnValleuArgGluThr--- 188
 DB 6417 AAAACACAGAGCTTGAAGTCTCATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6476
 OY 189 -----GlyAsnAsngluLeuValGlnleuThrglySerAspCysSerGlu 204
 DB 6477 GCAGAGGTGAAG 6533
 OY 205 SerAsnAlaGluilegluasleuSerGln---ValAspGlyProGluValGlnleu 223

```

Db 6534 CTGTTAAAGACAAACTATCTCCAGGAAAGACTGACAGTTTGGAAAGAGACTCAAG 6593
QY 224 LeuLeuSerThr-----ValGlnProAsnLeuGluLys 215
Db 6594 GCACCTGTTGCAAAATGTGAGTGGAAACCAATTGCACACTGAAATTAAGAGAA 6653
QY 236 GluValTrpGlyMetGluAsnSer-----SerGluSerPheAla 250
Db 6654 GAATTCCTGTTCAAGGATGTGAAGCCCTGCAGCCAGACTGAGTAAATTAATTAAG 6713
QY 251 AspSerValValSerGluSerAspThrSerLeuAlaGlySer----- 266
Db 6714 AACGTGATGCTCCAAAGCCCTGGAGGCCGACTGTTGAGAAAGTAAAGTTGCA 6773
QY 267 ----ValSerCysLeuAspGluSerLeuGlyHisAsnSerAsnMetCysAspSerGly 285
Db 6774 AGCCTGAGCTCAACACAGAGGGAAGTG---CATCACTGTGAAAGAGCAATGAGAA 6830
QY 286 ThrMetGlySerAspSerAspGlu-----GluAsnValAlaAla 298
Db 6831 AGAGTTCGCTTACAGCCGCTGAAAAAGACAGCTGCATCCAGAGAACTGAAAGAA 6890
QY 299 ArgAlaSerProGluProGluLeuGlnLeuArgProTyrGlnMetGluValAlaGlnPro 318
Db 6891 CCGGAGCGGAGAAATGCTTCACTTAAGGATTAAGTTGACAAACCTTGAAGAGAAAT 6950
QY 319 AlaLeuGluGlyLysAsnLeuLeuLeuLeuLeuProThrGlySerGlySerThrVal 338
Db 6951 ATGTGACAAACAAACAGAGAGTATGATCTTGCAGCCAGAAATTCACAGAGAGAA 7010
QY 339 AlaValTyrLeuAlaLys---AspHisLeuAspLysLysLysLysAlaSerProGly 357
Db 7011 GAACACTTAAACACAAATAGAAAGAGATGCCAGAGCTGAAAGATTTCGATTAGAC 7070
QY 358 LysValLeuLeuValAsnLysValLeuLeuValGluGlnLeuPheArgGluPhe 377
Db 7071 CTGTGACGTTAAGCTGTAAGAAAAAGAAATCTGCACAAATTCACAAAGAAAGAG 7130
QY 378 GlnProPheLeuLysLysTyrTyrArgValLeuGly-----LeuSerGly 392
Db 7131 CAG-----TTGCAGAACAGACAGAGTTACTCTCTCATTTAAAGTCTGTAAGAA 7184
QY 393 AspThrGlnLeuLysLysSerPheProGluValValLysSerCysAspLeuLeuSer 412
Db 7185 AAGGACCAACAGATACGATCAAGAGAGAAATCTAA-----ACT 7226
QY 413 ThrAlaGlnLeuGluAsnSerLeuLeuAsnLeuGluAsn----- 426
Db 7227 GCAGTGGAGATGCTTCAGATTCAGTTAAAGAGCTAAATAGCGAGTAAAGCTTCT 7286
QY 427 GlyGluAspAlaGlyValGlnLeuSerAspPheSerLeuLeuLeuLeuLeuGlyCysHis 446
Db 7287 GGTGACCAACAAATTATGAGGCCAGCAAGAGCTTACCCACCAATAGAGAGAGAG 7346
QY 447 HisThrAsnLysGluAlaValTyrAsnAsnLeuMetArgHisTyrLeuMetGlnLys 466
Db 7347 CATCAGCTGAGAAATACAT-----TAAAGATG 7376
QY 467 LysAsnAsnArgLeuLysLysGluAsnLysProValLeuProLeuProGlnLeu 484
Db 7377 AGAGCCCGCTAGAGCTGATGAAGAAAGACAGCTGTCTTACAAACAGCTGAAAGAA 7436
QY 484 ----- 484
Db 7437 AGTGAGCATGAGATTAATTACTTAAGGCTAGAGAGAACTTGAACAGAGCTAGAG 7496
QY 485 LeuGlyLeuThrAlaSerProGlyValGlyGlyAlaThrArgGlnAlaLysAlaGlu 504
Db 7497 ATACCCAGGACAAACAGAGCATGACGCTCTTGAGGACAGAAATTCACAAAGAGAGTA 7556
QY 505 HisLeuLeuLysLeu-----CysAlaAsnLeuAsp 514
Db 514 ----- 514

```

```

Db 7557 GAGACCCATAAGCAAAATATAGAGGATGACCCAAAGCTGAGAGTGTGAAATAGAT 7616
QY 515 AlaPheThrLeuLysThrValLysGluAsnLeuAspGlnLeuLysAsnGlnLeuGlu 534
Db 7617 GTTGTACTATAGGTCAGCAAAAGAAATTCGACAAATGAAATTAACAAAGAGAGAG 7676
QY 535 ProCysLysLysPheAlaLeuAlaAspAlaThrArgGluAspProPheLysGluLysLeu 554
Db 7677 CGAATATCTGAATTAAGAAATTAATTAATTCATTTGAAATATTTTGCAGCAATTA 7733
QY 555 LeuGluLeuMetThrArgLeuGlnThrTyrCysGlnMetSerProMetSerAspGly 574
Db 7734 -----GAGCAAGAAAGTACAGATTAAGAAATTAACAGTGTGCG 7775
QY 575 ThrGlnProTyrGluGlnTrpAlaLeuGlnMetGluLysLysAlaLysLysLys 594
Db 7776 ATGAGATGCTTCCAAACACAAATTAAGAGCCCAATGAGAGAGTGCAGCCCTGATAT 7835
QY 595 ArgLysGluArgValCysAla-----GluHisLeuArgLysTyrAsnGluAlaGln 612
Db 7836 GACCAAGAACCTGTAAGGCCAAAGACAGCAATTTAGTACCAAGTACAGTGTTCGA 7895
QY 613 IleAsnAspThrLeuArgMetLeuAspAlaTyrThrHisLeuGluThrPheTyrAsnGlu 632
Db 7896 CTT---GAGAGAGCTCAGTGTGACAAAGCCCTTGATGAGCCCAAAATATATATTTGT 7952
QY 633 GluLysAspLysLysPheAlaValLeuGluAspAspSerAsp-----Glu 647
Db 7953 TTGCATCTTCAGTGAAGAGCCCTCATTCAGAAAGTAAAGATGAGCAAGAGAGAA 8012
QY 648 GlyGlyAspAspGluTyrCys-----AspGlyAspGluAspLysPheLys 664
Db 8013 AAGACAGATGAACAAATCTGAGCTGAGAAATCAAAATTCAGACCAAGAGAGAGAG 8072
QY 665 LysProLeuLysLeuAspLysThrAspArgPheLeuMetThrLeuPheMetGluAsn 684
Db 8073 TCTAACTGCTCCAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8123
QY 685 LysMetLeuLysArgLeuAlaGluAsnProGluTyr-----Glu 697
Db 8124 TTAGAACTAGAAATCTGACGCTGGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8183
QY 698 AsnGluLysLeuThrLysLeuArgAsnThrLeuMetGluGlnTyrThrArgTrpGlu 717
Db 8184 AATGCTCTTTCAGAGACACATTAAGAGCTGAGAGAGTCTTACAAAGATTAAGAGAT 8243
QY 718 SerAlaArgGlyLeuLeuPheThrLysThrArgGlnSerAlaTyrAlaLeuSer 737
Db 8244 GAG-----CTGAATTCAGCAAAATGAGCAAAATGCTCTTCTT 8282
QY 738 IleThrGluAsnGluLysPheAlaGluValGlyValLysAlaHisLeuLeuAla 757
Db 8283 -----CAAAATTAACAAATGAGCTGCAAGAAAGAGAGAGAGAGAGAGAGAGAG 8330
QY 758 GlyHisSerSerGluPheLysProMetThrGlnAsnGluGlnLeuValLysLys 777
Db 8331 ATGCATGACATGCGACAGAGAAACAGCAGAGCTGCCAAGAAAGAGCTAGTGAAATTA 8390
QY 778 PheArgThrGlyLysIleAsnLeu----- 786
Db 8391 AGCCTAGCTGAGAGTGTGAGTACTGTTGGAAGAAATTAAGAGAGAGAGAGAGAGAG 8450
QY 787 -----IleAlaThrThrValAlaGluGlyLeuAspTrpLeuSerGluAsn 802
Db 8451 AAGGACCTACACTAGAAATATAGTGAATTAAGAAAGAGAGAGAGAGAGAGAGAGAG 8510
QY 803 IleValIleArgTyrGlyLeuValThrAsnGluIleAlaMetValGlnAlaArg 822
Db 8511 CAGCTGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8570
QY 823 AlaArgAlaAspGluSerThrTyrValLeuValAlaHisSerGlySerGlyValLeuGlu 842
Db 8571 GAGGTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 8630

```

QY 843 HIGLITHTVALSNAPRHEATGILVYMETMETYLISALALEHTSYVALGID 862
 DB 8631 ATCCAGACA-----TACCAGAGAAATTACTTTAAAGAA---GATGTCTCAGT 8678
 QY 863 ASMETLYSPROLU-----GILVYFALAHSLYS(I)ELAN 874
 DB 8679 TCACAGAGCTGAGATAGACCTTTAAAGCTAGTAAGACAGCTCAAAATATCATTC 8738
 QY 875 GILVLEUGLIMETGLINSEIEMETGLVLYSLYSMETLYSHLYS----- 889
 DB 8739 AAAGTACTACTCAG---ATTGTGAAGAAATTGAACAAACCAAGATGCAATATCAAAA 8795
 QY 890 -----ARGSNILEALALYSHTYRLYSASAPNPROSERLEULEHTPIELVUCYS 907
 DB 8796 TATGTAAATCAGTTGAAGAAAGAAATGAACGTGCCAGGGAATAATGTTGTGATC 8855
 QY 908 LYSANCSYSSERVALLEUACYSSEGLVLUASPILEHISVALLEGLIULYS----- 925
 DB 8856 AATCTCTGTAACAGCTGGA-----GAGGAAAGACATCTGTCAGCAAGAACATC 8906
 QY 926 -----METHISHIS 928
 DB 8907 TCTCACTTCAAGCTGCACAGAGAGAAACAGAGTACTGTATGAA ACCAAGTC 8966
 QY 929 VALASMETHTPROGLUPHELYSGILVLEUTYLILEVALARGULASNYLSALA----- 946
 DB 8967 GATGAATTAACACTGAGATCAAGAAAGAAAGAACTCTTCAAGAAACCAAGAGAG 9026
 QY 947 -----LEUGLULYSLYSCYSALEAPTYGLILEANSGLIGLILELLEYS 952
 DB 9027 GCAGATGAATCTGATAGTACTCTCTCC---TTCCTRTAAGCCATGAAGTTACAG 9083
 QY 963 LYSGLYGLINALATRPGLYTHRMETMETVALHIS 974
 DB 9084 AAAGCTAAAGAGATGTAGAGACACAAAGTGGCCAT 9119

RESULT 11
 US-08-685-576-5
 Sequence 5, Application US/08685576
 Patent No. 5906819
 GENERAL INFORMATION:
 APPLICANT: KALBUCHI, KOZO
 APPLICANT: IWAMATSU, AKIHIRO
 APPLICANT: NAKANO, TAKESHI
 APPLICANT: ITO, MASAOKI
 APPLICANT: TAKAHASHI, NO. 5906819AKI
 TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE
 NUMBER OF SEQUENCES: 16
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: FOLEY & LARDNER
 STREET: 3000 K Street, N.W., Suite 500
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20007-5109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.10
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/685,576
 FILING DATE: 24-JUL-1996
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 7-325129
 FILING DATE: 20-NOV-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 8-17150
 FILING DATE: 05-JAN-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 8-131206

FILING DATE: 26-APR-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Bent, Stephen A.
 REGISTRATION NUMBER: 29,768
 REFERENCE/DOCKET NUMBER: 16887/843
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202)672-5300
 TELEFAX: (202)672-5399
 TELEX: 904136
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 4363 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..4164
 US-08-685-576-5

Alignment Scores:
 Pred. No.: 1,71e-10
 Score: 188.50
 Percent Similarity: 35.43%
 Best Local Similarity: 18.51%
 Query Match: 3,55%
 DB: 2
 Gaps: 59

US-09-515-363c-2 (1-1025) x US-08-685-576-5 (1-4363)

QY 71 TRPHISLEUGLYTRPHTHARGSLUPHEVALGLUALALEUARGVARGTHGLYSERLEUEN 90
 DB 1075 TGGCAT-----TGG-----GATTAACATTAAGAAAGGAGCGACCTCTCTCA 1113
 QY 91 ALAALAARGTYMETASNPROGLUEN---THRASPLEUPROSERPROSERHTLUSN 109
 DB 1114 GTA-----CCTGAACTCAGCACTGACATAGACAGGAGCAATTTCATVAG 1158
 QY 110 ALAHISASP-----GLUTYLEU-GLINLEUENLEULEU----- 121
 DB 1159 ATTGAAGATGACAAAGAGATGTACAAACCTTCCCAATTCCTAAAGCTTTGT 1218
 QY 122 -----GLNPROTHLEUVALASPLYSLEULEUVALARGVSTVALLE 135
 DB 1219 CAGCTGCTTCATCGGATTTACCTACTATAGA-GAAATTTATATTATAGTGA---TC 1274
 QY 135 UASPLYSCTSMETGLUGLUGLULEUHTHTLLEGLUSPARGASNARGILEVALAL 155
 DB 1275 TCCATCTTTAGAGAAATGAT-----TCCATACATCAAGGAAA----- 1314
 QY 155 AGLUSNANGLYASNGLUSERGLYVALARGLULEULEUYSARGILEVALGLLYG 175
 DB 1315 -----AATGAAGAAAGT---CAAGACATTCAGAAAAAAGCTGTATAATTGA 1358
 QY 175 UASNTRPHESEALAPHELEUASNVALLLEUARGSLNTHRGLYASNANGLULACVVALG 195
 DB 1359 AGAACATCTTAC-----AATGAGATCGAAGCCAA 1388
 QY 195 NGLULEUTHRGLYSERASPCYSERGLUSERASNALAGLILEGLUANLEUSNGINVA 215
 DB 1389 ACAGCAACTGCAACAACTGCAATCTGTTAATCTCGCCCTAGCAAAAACAGAAAG- 1446
 QY 215 LASPLVPROGLINVALGLUGLUGLULEUUSERHTHTRYVALGINPROASNLGLULY 235
 DB 1447 -----GAGCTRAGAGAGAGAGATT-----ACCTTACGGAAGAGTGTAATC 1487
 QY 235 SGLUVALTRPGLYMETGLUASNSNSERSEGLUSERSEPHALASPSERSEVALVA 255
 DB 1488 ACCATTAAACAGACGTTAGAAAGAGAAAGGCGCTTCTTCAGCACAATAATGCANATATCA 1547

QY 255 lsergluserasphrserleu-----AluGluclyserValserCysle 270
 : : : : :
 Db 1548 GAGGAAGCTGATCATGAGACAGACAAAGAAATTTGAAATGATGCTTACAGCTT 1607
 QY 270 uaspgluserleu-----GlyHisAnserSermetCysSerAspse 284
 : : : : :
 Db 1608 AAAAGATCAACTGAGATTTGAAAAAAGAAATCAAAATCTCAAAATATACACTGAGAA 1667
 QY 284 r---GlyThrMetGlySerAspSerAspGluGluValAlaAlaArgAlaSerProgl 303
 : : : : :
 Db 1668 AGTGAATCAACTGACAGACAACTGATGTAACCAATGCTTACAGCGAAGAGCTGCA 1727
 QY 303 uprogluleuGlnleuArgProtyrGlnmetGluValAlaGlnProala----- 319
 : : : : :
 Db 1728 TACTGACGCCGCTTAAGAGAAACCCAGACAGAAAGTTCAAAACATTCACAGCTGCA 1787
 QY 320 -----LeuGluGlyAsnIleIleIleCysLeuProThrGlySerG1 334
 : : : : :
 Db 1788 ATCTAACATAGAGATCTACAGATTAAGATCAAGCTGCTG-----CTGAGACATGCCAAGTT 1841
 QY 334 ylystrhratrgValAlaValTyrIleAlaIlyAspHisLeuAspLysLysLysAla 354
 : : : : :
 Db 1842 AAAACTTGAAGAAAGCAATTTATCAATCTTCACTGACCTCTAGATCTGAAACAGGATCG 1901
 QY 354 rgluproglLysValIleValleuValAsnLysValleuValGluGlnleuPheAr 374
 : : : : :
 Db 1902 AACCATGATCGATCGATATATT----- 1923
 QY 374 glyslupheGlnProPheleuLysLysTyrPheArgValIleGlyLeuSerGlyAspTh 394
 : : : : :
 Db 1924 -AATGATTTACAGAGT-----AGAACATGCGCTTAAGACAT-- 1962
 QY 394 rgluleuLysIleSerPheProgluValValLysSerCysAspIleIleIleSerThrAl 414
 : : : : :
 Db 1963 -----TTAAAGACGGTAAATCTTATACCGAAGT 1994
 QY 414 aglnIle-----LeuGluAsnSerLeuLeuAsnLeuGluAsnIlyGluAspAl 430
 : : : : :
 Db 1995 AGAAGTGAAGACAGACAACTTCAAGATTCAGAGATTCTGATTTGAAAG--GAAAGAA 2051
 QY 430 aglyValGlnleuSer---AspPheSerleuIleIleIle-----As 443
 : : : : :
 Db 2052 CAACATGGAAGATGATGATGATGATCAACCACTAAAGTTATACACAGAGCTTACACAGCA 2111
 QY 443 pglucySHSHIS-----ThrAsnLysGluAlaValIlyAsnAs 456
 : : : : :
 Db 2112 AGAAGCTGAAGACATGAAGCCCAAGACAGACACTAGACAGATAAATAATGACA CTATGCTC 2171
 QY 456 nileMetArgHisTyrIleuMetGlnLysLeuLysAsnArgLeuLysIlyGluAsnLys 476
 : : : : :
 Db 2172 CATC-----GAAAGACCAATATCGAAGCCATGAAACAAATGAGGAA 2213
 QY 476 sprovalIleProleuProGlnIleleuGlyLeuThrAlaSerProglValGlyAla 496
 : : : : :
 Db 2214 GAAGCTCTTGGAG-----GAAG 2231
 QY 496 aThrLysGlnAlaLysAlaGluGlnHisIleleu-----LysLeuLysAlaAsnLe 513
 : : : : :
 Db 2232 AACTTTAAACAGAAAGTGAAGAACTATTGCTAGAACTGAGAAAGATTTCTCTATT 2291
 QY 513 uasPalaPheThrIleLysThrValLysGluAsnLysAspGlnleuLysHisGlnIleG1 533
 : : : : :
 Db 2292 AGAGCTG-----GACCTCAACAGCTGACAGAGAGAGAAATAAA 2327
 QY 533 ngluprocysLysLysPheAlaIleAlaAspAlaThrArgGluAspProLysGluLys 553
 : : : : :
 Db 2328 TGAGCTCTTAACAG-----AAAGATGCTGCTAAATGAGGAT--TAGAAACCT 2375
 QY 553 sleuLeuGlnIleMetThrArgTyrGlnIleThrTyrCysGlnMetSerProLysSerPhe 573
 : : : : :
 Db 2376 GACATTAATAATGAGACAGAAACTGAGAGCGCTC-----CTTACATTAATGAGCTT 2429
 QY 573 eglYThrGlnProtyrGluGlnIleThrAlaIleGlnMetGluLysLysAlaLysLysG1 593

Db 2430 GAAGATGCAAAACACAAAGGTTAACACACTAAAAATGTCAGAAAAAGCAATTAAAGCAGA 2489
 : : : : :
 QY 593 yAsnArgLysGluValGlyAlaGlnHisLeuArgLysTyrHisGlnValleu----- 611
 : : : : :
 Db 2490 AATTAACCATCTCATGAAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGA 2549
 QY 612 -----GlnIleAsnAspThrIleArgMetIleAspAlaIlyThr 624
 : : : : :
 Db 2550 AGAAGCTGAGATGATGATGAGTGGCAAAATGAAAGAGCTCCAGATGACCTGAAAGCAGACA 2609
 QY 624 rHisLeuGlnThrPheTyrAsn-----GluGluLys 634
 : : : : :
 Db 2610 GTATTCTCAACCTTTATTAACACAAAGTTAGGAGCTTAAGCAAGATGAGAGAA 2669
 QY 634 sasLys-----LysPheAlaValIleGluAspAspSerAsp-- 646
 : : : : :
 Db 2670 GACCAAACTTGTTAAAGATTGACAGCAGAAAGAACAGAAATTAACAGATGAAAGCGACTC 2729
 QY 647 -----G1 647
 Db 2730 TTGGCTGCCCACTGAGATGATGATGACCAAGACATTTCTGACAACTGCTTCTG 2789
 QY 647 uGlyLysAspAspGluTyrCysAspGlyAspGluAsp-----GluAspAspLe 663
 : : : : :
 Db 2790 AATGCTGAAGAAACAATATTCTGATTTGAAAAAGAAAGATCATGAAGACTGACAT 2849
 QY 663 ulysLysProleu-----LysLeuAspGlnThrAspArgPheLeuThr 678
 : : : : :
 Db 2850 CAAGAGATGATGAGCTGACAGCAAAACAGAACTTACGGAAGAAAGATGTACAAATTCTC 2909
 QY 678 leuPhePheGluAsnAsnLysMetLeuLysArgLeuAlaGluAsnProgluTyrGluAs 698
 : : : : :
 Db 2910 TCTT---GAGGAACCTAATAGACACTACTAGTGTGTCATCTTGCAATGACAA 2966
 QY 698 nglLysLeuThrLysLeuArgAsnThrIleMetGlnIlyThrArgThrGluGlu-- 717
 : : : : :
 Db 2967 ACAGACATTAATAATCAAAATTAAGATGTTGCAAGCAACTGCAAGATTTGAAGATGA 3026
 QY 718 -----SerAlaArgGlyIleIlePheThrLysThrArgIleAsnAlaIlyLeu 735
 : : : : :
 Db 3027 AGAATTAACCGACGACCTATT-----MAAGCACTTGACAGA 3065
 QY 735 rgluprIleThrGluAsn-----GluLysPheAlaGluValG1 748
 : : : : :
 Db 3066 GCAGCTATTACAGAAAGAACACTCAAACTCAAGCTGATGATGATGATGATGATGATGATG 3123
 QY 748 yValLysAlaHisLysLeuIleGlyAlaGlnHisSerSerGluPheLysProMetIlyG1 768
 : : : : :
 Db 3124 -ATGATCGAAAAAGAACTGTCAGACCTGCTATATGACAGAGATGTGCG--AGAAAGA 3179
 QY 768 nasngLuglnLysGluValIleSerLysPheArgThrGlyLysIleAsnLeu----- 786
 : : : : :
 Db 3180 GAAGGAGAAATGAAAGCTACATATGAGCTTAATGTCAAGCTGACAAATGAC--AGCA 3239
 QY 787 -----IleAlaThrThrValAlaGluGluLysLe 796
 : : : : :
 Db 3240 GATGATCGAAGTATGACAAAGAACTGAATGAATGACAGGACAAATAGCTGAGAGA-- 3294
 QY 796 uasrIleLysGluCysAsnIleValIleArgTyrGlyLeuValThrAsnGlnIlyLeu 816
 : : : : :
 Db 3295 -----AGCGAATTCGAATTAAGTCAAGATGATGATGATGATGATGATGATGATGAT 3347
 QY 816 tValGlnAlaArgGlyArgAlaArgAla-----AspGluSerThrThrVal 831
 : : : : :
 Db 3348 TGAGCAGCTGCGCTGACAACTGCAAGCTTGCATATGCTGCTGCTGCTGCTGCTGCTGCT 3405
 QY 831 lleuValAlaHisSerGlySerGlyValIleGlnHisGluThrValAsnAspThr 851
 : : : : :
 Db 3406 -----GCGAGTGGACAGGAGATGCTGACAGATGATGATGATGATGATGATGATGAT 3449
 QY 851 ulysMetMet-----TyrLysAlaIleHisCysValGlnAsnMetLysProLysPhe 869
 : : : : :

```

Db 3450 ATCAAGATTAGAGATGCTTTCATTCCTGCTAGCAAAACACCTAAGCTAATTCGATG 3509
QY 869 TAlHILyStIleuGluuGlnMetGlnSerIleMeGluLysLysLeuLysThrLy 889
Db 3510 GGTAAAGATATGTGATTGTAAGCACTATCTTTCTATACACAGTGAACAGCA 3569
QY 889 sArGAsnIleAlaLysHISLySAsnAsnProSerLeuIleThrPheGluCysLysAs 909
Db 3570 TAAAGAA-----CATCCAACTCTTACATGCTTTAGATTAAGACAAATT 3614
QY 909 nCySeSerValleuAlaCySerGlyGluAspIleHisValIleGluLysMetHisIleVa 929
Db 3615 ATTTCATGTCGACCACTTACACAGACAGACAGATGCTAT----- 3651
QY 929 lAsnMetThrProGluPheLysGluLeu-----TyrIleValArgGluAsnLy 945
Db 3652 -----AGACAGATCTTAAAGAAATTCACAGATATTCAGATTCGTATGCCAACTGA 3704
QY 945 sAlaLeuGluLysLysCysAlaAspTyrGlnIleAsn-----LysGlu----- 3760
Db 3705 AGGAGAAAGTAAAGAGACAAACAAATTCACAGTGAAGCCAGTGGAGAAATCTAATTA 3764
QY 960 eIleCyLysCyGlyGlnAlaIlePheLysThrMetMetValHis----- 3764
Db 3765 TATTGCCACAGGACATGAGTTTATTCTACTACTTATCATTTCCCAACCACTGCGA 3824
QY 974 ----- 3824
Db 3825 GCGTTGATGAGCCCTGTGGACATGTTAAGCTCTCTCTCTTGTAGTGGCCGCG 3884
QY 975 -----LysGluLysLeuPheProGly 981
Db 3885 TTGGCATATTAAAGTATCAATGATATGACAAAGAGAGACATTTACACACTTCG 3944
QY 981 sIleuLysIleArgAsnPheValValPheLysAsn-----AsnSerThr 996
Db 3945 C---AAAGTATTATTCATTTTTCACGCAAGAAATCTGTATTACTAAATTTAC 4001
QY 996 IlyLysGlnIleLysLysThrVal 1004
Db 4002 AGAAGACAG-----CAGAGTGGCTT 4023

RESULT 12
US-09-643-597-117
: Sequence 117, Application US/09643597
: Patent No. 6426072
: GENERAL INFORMATION:
: APPLICANT: Wang, Tongtong
: APPLICANT: Fan, Liqun
: APPLICANT: Kalos, Michael D.
: APPLICANT: Banquer, Chaitanya S.
: APPLICANT: Hosken, Nancy
: APPLICANT: Fanger, Gary R.
: APPLICANT: Li, Samuel X.
: APPLICANT: Wang, Aijun
: APPLICANT: Skelky, Yasir A.W.
: APPLICANT: Henderson, Robert A.
: APPLICANT: McNeill, Patricia D.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE TREATMENT
: OF INHERITED AND ACQUIRED DEFECTS OF LONG CANCER
: FILE REFERENCE: 210121.455C11
: CURRENT APPLICATION NUMBER: US/09/643,597
: NUMBER OF SEQ ID NOS: 369
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 117
: LENGTH: 6921
: TYPE: DNA
: ORGANISM: Homo sapien
US-09-643-597-117
Alignment Scores: 5.05e-10 Length: 6921

```

```

Score: 187.50 Matches: 214
Percent Similarity: 34.688 Conservative: 164
Best Local Similarity: 19.638 Mismatches: 375
Query Match: 3.538 Indels: 337
DB: 4 Gaps: 51

US-09-515-363c-2 (1-1025) x US-09-643-597-117 (1-6921)
QY 41 GluValIleGluGlnIleGlnArgThrValAlaThrSerGlyAsnMetGlnValGlu 60
Db 1313 GAAATACAGCTGAAATATGTAATTAACCTAGCCACAGCACTTGAATCAACAG-----AAG 1366
QY 61 LeuLeuLeuSerThrLeuGlnLysGlyValIlePheIleGlyIleThrArgGlnIleVal 80
Db 1367 ATGCTGTGTCGCAATTAAGAAATGAACAGACCAAAATGACAGAGTGTCAAAATATGCA 1426
QY 81 GluAlaLeuArgArgThr----- 86
Db 1427 GAACACTACTACAGCTACAGTGAAGCACTGATTAACAAACATGACCTACGCAACATG 1486
QY 87 -----GlySerProLeuAlaAlaArgTyrMetAsnProGluLeuThrAsp 1501
Db 1487 GTAGATTCAACAAACAAATCTCAGTGAACCCGCAAGATGCAAGTCA-----GCAAT 1543
QY 102 LeuProSerProSerPheGluAsnAlaHisAspGluIleGluLeuLeuAsnLeuLeu 121
Db 1544 CTCATTATTCAAGAGCTTCAAGCTGAGCTTACAGCTGATCTGCTGCTCAATTCAGT 1603
QY 122 GlnProThrLeuValAspLysLeuValArgAspValLeuAspLysCysMetGlnIle 141
Db 1604 ACACATATAT-----AAATTGCTGTGATTCATTAAAGCTGCAACAGGAG 1654
QY 142 GluLeuLeuThrIleGluAsp-----Arg 149
Db 1655 GAGATTAAAGGTGTAGAGAGAGACTTCTGACATGGGGCATATTCAGATCTGCTTACGGT 1714
QY 150 AsnArgIleAlaAlaIleGluLysAsn-----GlyAsnGluSerGlyValIleGln 166
Db 1715 CAGAGGCAACAGCTGCTGAGATGACAACTTACAGCAAGACATAGTAGAGTGAAGA 1774
QY 167 LeuLeuLysArgIleValGlnLysGlnLysThrPheSerAlaPheLeuAsnValIleVal 186
Db 1775 ATGCTAGCTGAACCTAAGCAACAAAGTCCGAGTAGAGAAAGACTTCCGAAAGTCCAG 1834
QY 187 GlnThrGlyAsnAsnGluLeu-----ValGlnGluLeuThr----- 198
Db 1835 GAGCTGCAGAAATGAATGATGAGAAAGCAGACAGAAATGTAGAGATATCTGCTGAG 1894
QY 199 -----GlySerAspLysSerGluSerAsnAlaGluIleGluLysMetSer 213
Db 1895 AAGATTAAAGGCTGAAAGTGAAGCCAAAGCAAGTACCCAGAGGAATGGAACCTTTGAGA 1954
QY 214 GlnValAspGlyProGlnValGlu-----GluGlnLeuLeuSerThrThrValLeuThrAsn 232
Db 1955 GAGAGAGAAAGCCGCTGAAGAGAACTGAGCGCGTAGAGCAAGCTCAGCATAGAATGCGAG 2014
QY 233 -----LeuGluLysGluValIlePheGlyMetGluAsnAsnSerSerGluSer 247
Db 2015 GCTAAAGAGCTGCGGTGAAGAGAACTCTGAAATTTCCCAATCACTGGAACCAAAAC 2074
QY 248 SerPheAlaAspSerSerVal-----ValSerGluSerAspThrSerValAlaIle 264
Db 2075 ACCTTTACAGAGCAACACTGGAACATCATCTTAAACAAAGATTTAAGTCT----- 2128
QY 265 GlySerValSerCysLeuAspGluSerLeuGlnLysAsnSerAsnMetGlySerAspSer 284
Db 2129 -----AATGATTTGGAGCAACAAATAATTAATTAATTAATTAATTAATTAAT 2167
QY 285 GlyThrMetGlySerAspSerAspGluLysValAlaAlaArgAlaSerThrArgPro 304
Db 2168 TTAAGAGAAAGAGAGAGCAATGAGAGAACTCTGAGAGTGTATTAACAGATGAGAGAAA 2227
QY 305 GluLeuGlnLeuArgProTyrGlnMetGluValAlaGlnProAlaLeuGlnLysLysAsn 324

```


QY 921 HisValIleGluLysMetHisHisValAsnMetThrProGluPheLysGlnLeuTyrIle 940
 Db 4022 CTGTCTCTAGAAAGACTGGACACTTGCACCCAGACAGCCAGATCTCCCTGTTGAGATGG 4081
 QY 941 ValATGCTGAsnLysAlaLeuGlnLysLys 950
 Db 4082 ACTCAAGAACACACAGCATTTGGAAGACAG 4111
 RESULT 13
 US-08-961-527-142
 ; Sequence 142, Application US/08961527
 ; Patent No. 6420135
 ; GENERAL INFORMATION:
 ; APPLICANT: Charles Kunsch
 ; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
 ; NUMBER OF SEQUENCES: 391
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
 ; COMPUTER: HP Vectra 486/33
 ; OPERATING SYSTEM: MSDOS version 6.2
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/961,527
 ; FILING DATE:
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Brookes, A. Anders
 ; REGISTRATION NUMBER: 36,373
 ; REFERENCE/DOCKET NUMBER: PB340P1
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 309-8512
 ; TELEFAX: (301) 309-8504
 ; INFORMATION FOR SEQ. ID NO: 142:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 5020 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; US-08-961-527-142
 Alignment Scores:
 Pred. No.: 3,26e-10 Length: 5020
 Score: 187.00 Matches: 157
 Percent Similarity: 34.53% Conservative: 132
 Best Local Similarity: 18.76% Mismatches: 263
 Query Match: 3,528 Indels: 285
 Db: 4 Gaps: 39
 US-09-515-363c-2 (1-1025) x US-08-961-527-142 (1-5020)
 QY 309 ATGProTyrGlnMetGluValAlaGlnProAlaLeuGlnLys-----LysAsnIleIle 326
 Db 1097 CAAGCCCAAGCTATGAGAGAGTTGGTGGATTAACATTGAGGGGAGAAAGACCTCAGATT 1156
 QY 327 IleCysLeuProThrGlySerGlyLysThrArgValAlaValTyrIleAlaLysAspHis 346
 Db 1157 CTGATGGGGGGCGACTGGAGACGAGAGACC-----TATACTATGAGTCCAGCTC 1204
 QY 347 LeuAspLysLysLysLysAlaSerGluProGlyLysValIleValLeuValAsnLysVal 366
 Db 1205 ATTCTAAAGTCAATAAACCAACT-----CTGTATTGCTCCACAAATAA 1249

QY 367 LeuLeuValGlnGlnLeuPheArg-----LysGluPheGlnProPheLeuLys, sTrp 384
 Db 1250 ACTGTGGCTGTGTCAGCTCTATGGGAGATTAAAGAAATTTTCCCT----- 1294
 QY 385 TyrArgValIleGlyLeuSerGlyAspThrGlnLeuLysIleSerPhe----- 400
 Db 1295 -----GAAATGCAGTTGAGATATTTGCTTCTTACTATCATTAATTC 1336
 QY 401 ---ProGluVal---ValLysSerCysAspIleIleIleSerThrAlaGlnIleLeuGlu 418
 Db 1337 CAGCCAGAGCCCTATGCTCCCTCTAGCCATACCTATATGACAGAGATAGTCTTCATAT 1396
 QY 419 AsnSerLeuLeuAsnLeu---GluAsnGlyGluAspAlaGlyValGlnLeuSerAspPhe 437
 Db 1397 GACGAGATTGACAAAGCTTGCCACGCTACCTACCTAGCCCTTTTGAGAGCTATATATGTT 1456
 QY 438 SerLeuIleIleIleAspLysCysHisHis-----ThrAsnLysGluValValTyr 454
 Db 1457 ATTTGCTGGCTCAGTCTTGTATCTATGTTGGTTGGCTCCCAAGAAATACCTGAT 1516
 QY 455 Asn-----AsnIleMetArgHisTyrLeuMetGlnLysLeu 466
 Db 1517 AGTGTGCTAGTCTGCGCTGCTGCTAGAGATTCTCGTGAATTAACCTTGATATGATTGG 1576
 QY 467 -----LysAsnAsn-----ArgLeuLys 472
 Db 1577 GTGATATTACCTATTGAACGTAATGATATTGATTTCACAGCGGAAGATTTCGCTTCGT 1636
 QY 473 LysGluAsnLysProValIleProLeuProGlnIleLeuGlyLeuThrAlaSerProGly 492
 Db 1637 GGGGATGTGTAGACATTTCCTCA----- 1660
 QY 493 ValGlyValAlaThrLysGlnAlaLysAlaGlnGlnHisIleLeuLysLeuCysAlaAsn 512
 Db 1661 -----GCTTCCGAGATCAACAT----- 1678
 QY 513 LeuAspAlaPheThrIleLysThrValLysGluAsnLeuAspGlnLeuLys----- 529
 Db 1679 -----GCCTTTCGAGTGAATTTTGTGAGACGAATTTGCTGTAATTCGTAATTCGAG 1732
 QY 530 -----AsnGlnIleGlnLysProCysLysPheAlaIleAlaAspAla----- 544
 Db 1733 GCTGTGACAGGTGACGTGTGGAGACAAGTGATCTTTAGCATTTTCCCAATGATACAC 1792
 QY 545 -----ThrArgGluAspProPheLysGluLysLeuLeuGlnIleMetThrArgIleGln 562
 Db 1793 TTGTGACCAATGACGACACATGAGATGTCATTTGCAAAATTCACAGCCAGATTCGAA 1852
 QY 563 ThrTyrCysGlnMetSerProMetSerAspPheGlyThrGlnProTyrGlnGlnIleAla 582
 Db 1853 -----GAAATTAATGCT 1864
 QY 583 Ile---GlnMetGluLysLysAlaIleAlaLysLysGlyAsnArgLysGluArgVal----- 599
 Db 1865 GTCTTTGAAAGAGAGATTAAGCTGTTGAGCCCAAGCTTTGAAACAGCGGACAAATAT 1924
 QY 600 CysAlaGlnHisLeuArgLysTyrAsnGluAlaLeuGlnIleAsnAspThrIleLeuMet 619
 Db 1925 CATATGCAAAATGCTCGCGACATGCGGTATACCAATGAGGGTTGAAATATATCTGTCAC 1984
 QY 620 IleAspAla-----TyrThrHisLeuGlnIleThrPheLysAsnGlu 632
 Db 1985 ATGATGAGACGAGGAGGAGAGAGAGCCCTCTTATACGCTTTCGACTTCCTCAATGAT 2044
 QY 633 GluLysAspLysLysPheAlaValIleGluAspAspSerAspGlu-----GlyLysAsp 650
 Db 2045 -----TTCTGATTTATGATTGACGAGATCATATGACATAGAGAAATC 2089
 QY 651 AspGluTyrCysAspGlyAspGluAspGluAspAspLeu----- 663
 Db 2090 AAGGCGATGTAACATGAGACCGTCCGGTAAAGAAATGCGTGAATTATATGTTTCCT 2149
 QY 664 -----LysLysProLeuLysLeuAspGluThrAspArgIleGlnMet 677

```

Db 2150 TTGGCGCTGCTTGGACAATGCTCTCCGCGGAGAGAGTGGAGAT GACGTTGAT 2209
QY 678 ThrLeuPheGluAsnLysMetLeuLysArgLeuAlaGluAsnProGluTyrGlu 697
Db 2210 CAGATGTTTACCTTCA-----GCCGACCTGTTGCTATGCA 2248
QY 698 AsnGluLysLeuThrLysLeuArgAsnThrIleMetGluGlnTyr----- 712
Db 2249 AATGAACAG-----ACCGAGACAGTATGACCAATACTCTCCCAACGGCA 2296
QY 712 ----- 712
Db 2297 CTCTGGATCCAGAGGTGGAAGTCCCTCCGACTATGAGACAGATGATGACTCTTGCT 2356
QY 713 -----ThrArgThrGluGluSerAlaArgGlyIleLeuPheThrLysThrArgGlu 729
Db 2357 GAATGATATGCGCGCTTAAAAAATGAGAGTACTTATCAAACTTGGACCAAGAA 2416
QY 730 SerAlaTyrAlaLeuSerGlnTrpIleThrGluAsnGluLysPheAlaGluVal 749
Db 2417 ATGGCAGAGATTGACCCAGTAC-----TTCAAGGAAATGGGTATC 2458
QY 750 LysAlaHisLysLeuIleGluAlaGlnHisSerSerGluPheLysProMetThrLysAsn 769
Db 2459 AAGGTCAAGTACATG-----CACTCGATATCAAGACCTTGG----- 2494
QY 770 GluGlnLysGluValIleSerLysPheArgThrGlyLysIleAsnLeuAlaIleAlaThr 789
Db 2495 GAACGAGCAGAGATTATCCGTCAGCTCGCTGGTGTGTGTGATGCTGTGTGGGAAT 2554
QY 790 ThrValAlaGluGluLysLeuAspIleLysGluCysAsnIleValIle----- 805
Db 2555 AACCTGCTCCGTGAAGAAATGACCTTCCGAAAGTGAAGCTGCTACTATCTCGATGCT 2614
QY 806 ---ArgTyrGluLeuValThrAsnGluIleAlaMetValGlnAlaArgGlyArgAlaArg 824
Db 2615 GACAGAGAGGCTTCCCTCGCAAGAACGTGACATCCAGACACATGACCTGCTGCA 2674
QY 825 AlaAspGluSerThrTyrValLeu-----ValAlaHisSerGlySerGly 839
Db 2675 CGTAATACCGCAAGGTATGATCATGATGATGCGGACAGCGTAAACAGTATGCAAGCT 2734
QY 840 ValIleGluHisGlnThrValAsnAspPheArgGluLysMetLeuLysAlaIleHis 859
Db 2735 GCTATGCAAT---GAAACTGCGCCCGCTCGCAAAATCCAGATGCTATTAAAGAAACAT 2791
QY 860 CysValGlnAsnMetLysProGluGluTyrAlaHisLysIleLeuGluLeuGlnMet 878
Db 2792 ---CGTATCGTTCACAAACCAACATCAAGAAAGAAATCCGTGACTTATTGCTGTC 2842
QY 879 ---GlnSerIleMetGluLysLysMetLysThrLysAspAsnIleAlaLysHisTyr 897
Db 2843 ACCAAGCGCACTGCTAAGGAGAGAACAGCAAGAGTGCATATCAATGCCCTGAACAAACA 2902
QY 897 AsnAsnProSerLeuIleThrPheLeuLysLysAsnCysSerValLeuLysSerGln 917
Db 2903 GAGC-----GCCAAGAACTACTCA----- 2921
QY 917 YguAspIleHisValIleGluLysMetHisValAsnMetThrProIlePheLysGln 937
Db 2921 ----- 2921
QY 937 uLeuTyrIleValArgGluAsnLysAlaLeuGlnLysCysAlaAspIle----- 954
Db 2922 -----AAAGCTTGAGAAACAAATGCACAGAGAGTCTTGAGTG 2959
QY 955 -----GlnIleAsnGluGluIleLeuLysCysGlyValIleAlaIlePro 968
Db 2960 CTGACTTTGAAGTACAGAGCTGAGATTGCTGATATGATGCTGAAGTCAAAAGCTTGAT 3019
QY 969 -----G 969

```

```

Db 3020 TAGGGAAATAGTATGATTTATTATTAAGAAAGTTAAGAAAGATTGATGTCIATGC 3079
QY 969 yThrMetValHis-----LysGlyLeuAspLeuProCysLeuLys 983
Db 3080 GAAATGCGCTTATTCACACCTTATTCACAGCTTGGAAGAAAGTATGATGCTCCCATATGAT 3139
QY 983 sIleArgAsnPheValValPheLysAsnAsnSerThrLysGln 999
Db 3140 GATATCAGTATTTTCAAAATTTTAAAGAAATTCGAATTCGAATAAATCAG 3188

RESULT 14
US-08-973-462-2
; Sequence 2, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DROULHE, PIERRE
; APPLICANT: DOUBRESTES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973.462B
; EARLIER FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 5361
; TYPE: DNA
; ORGANISM: P. falciparum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(5361)
US-08-973-462-2

Alignment Scores:
Pred. No.: 4,176-10 Length: 5361
Score: 186.50 Matches: 219
Percent Similarity: 35.09% Conservative: 181
Best Local Similarity: 19.21% Mismatches: 390
Query Match: 3,518 Indels: 351
DB: 4 Gaps: 49

US-09-515-363c-2 (1-1025) x US-08-973-462-2 (1-5361)
QY 42 ValLysGluGlnIleGlnArgThrValAlaThrSerGlyAsnMetGlnAlaValIleLeu 61
Db 967 GTTGAAGAAAGTGTAGCTGAAATGTTCAAGAAAGTGTAGCTGAAATGTTGAAATG 1026
QY 62 LeuLeuSerThrLeuGluLysGlyValIleThrPheLysGluGlyTrpThrArgGluIleValGlu 81
Db 1027 GTAGCTCAAGCTGTTCAAATAATCGTA----- 1053
QY 82 AlaLeuArgArgThrGlySerProLeuAlaAlaArgTyrMetAsnProGluLeuLeuAsp 101
Db 1054 -----GCTCCAACTGTTGAAGAAATGTAGCTGCAAGTGTGTTGAGAA 1095
QY 102 LeuProSerProSerPheGluAsnAlaHisAspGluTyrLeuGln----- 116
Db 1096 ACTGTGCTCCCAAGTGTGAGAAAGTGTGAGAAAGTGTGAGAAAGTGTGAGAAAGTGTG 1155
QY 116 ----- 116
Db 1156 AATGTGAAAGAAAGTGTAGCTGAAATGTTGAGAAAGTGTAGCTGAAATGTTGAGAA 1215
QY 117 -----LeuLeuAsnLeuLeuGlnIlePro 123
Db 1216 AGGTAGCTGAAATGTTAAGAAAGTGTAGCTGAAATGTTGAGAAAGTGTAGCTGCA 1275
QY 124 ThrLeu-----ValAspLysLeuLeuValArgAspValIleAsp 136
Db 1276 ACTGTGAAAGAAATGTAGCTCCAACTGTGAAAGAAATGTAGCTCCAAAGTGTGAGAA 1335

```

137 LysCysMetGluGluLeuThrIleGluAspArgAsnArgIleAlaIleAlaGlu 156
1336 AGTGGCTCCAACTGTTGAAGAAAGTGTGAAGAAATTTGAGAAAGAGTGTGAA 1495
157 AsnAsnGlyAsnGlyValArgIleLeuLeuLysArgIleValGluLysGluAsn 146
1396 AAT-----GTTGAAGAAAGTGTGAAAGAAATTTGAAAGAAATTTGAAAG 1449
177 ThrPheSerAlaPheLeuAsnValLeuArgIleThrLysAsnGluLysValGlu 146
1450 GAACAAAGTGTAGCTGAAGAAATTTGAAAGAAAGTGTAGCTGAAGAAATTT 1403
197 LeuThrGlySerAspCysSerGluSerAsnAlaGluIleGluAsnLeuSerGluValAsp 216
1504 AGTACTACTGAAATTTGTAAGAAAGTGTAGCT-----GAAATTTGTAAGAAATTTGTA 1557
217 GlyProGlnValGluGluLeuLeuSerThr----- 227
1558 GCTCAACTGTGTAAGAAATTTGTAAGAAAGTGTGCAAGAAATTTGTAAGTGT 1617
228 -----ThrValGlnProAsnLeuGluValIleProLysGluAsnSerSer 245
1618 GTAGAAAGTGTGCTCCAACTGTTGAAGAAAGTGT-----GAAAGAAATTTGTA 1668
246 GluSerSerPheAlaAspSerSerValSerGluSerAspThrSerIleValGluLys 265
1669 GAA-----AGTGTAGCTGAAGAAATTTGTAAGAAAGTGTAGCTGAAGAAAT 1710
266 SerValSerCysLeuAspGluSerLeuGlyLysAsnSerAsnMetGlySerAspSerGly 285
1711 -----GTTGAAGAAAGTGTAGCTGAAGAAATTT----- 1737
286 ThrMetGlySerAspSerAspGluAsnValAlaAlaArgAlaSerProGluProGlu 305
1738 -----GAAAGAAAGTGTAGCTGAAGAAATTTGTAAGAAATTTGTAAGTGTGAA 1788
306 LeuGlnLeuArgProTyrGlnMetGluValAlaGlnProAlaLeuGluLysAsnIle 325
1789 GAATTCCTACTCCAACTGTTGAAGAAATTTGTAAGAAATTTGTAAGTGTGTT 1836
326 IleIleCysLeuProThr-----GlySerGlyLysThrArgAlaValIleValIleAla 343
1837 GAAAGTGTGCTCCAACTGTTGAAGAAAGTGTGAAGAAATTTGTAAGAAATTTGTAAGTGT 1896
344 LysAspHisIleuAspLysLysLysAlaSerGluProGluLysValIleValLeuVal 363
1897 GAA-----AATGTTGAAGAAAGTGTAGCTGAAGAAATTTGTAAGAAATTTGTAAGTGTGAA 1947
364 AsnLysValLeuLeuValGluGlnLeuPheArgLysGluPheGlnProPheLeuLys 383
1948 AATGTTGAAGAAATTTGTAAGAAATTTGTAAGAAATTTGTAAGTGTGAA 2007
383 s-----TrpTyrArgValIleGlyLysSerGlyAspThr 394
2008 ATTGACTCCCAAGTGTGTAGAAAGTGTGCTCCCAAGTGTGAAAGAAATTTGTAAGAA 2066
394 ArgLeuLysIleSerPheProGluValIleLysSerCysAspIleIleIleSerThrAla 414
2067 AATGTTGAAGAAAGTGTAGCTGAAGAAATTTGTAAGAAATTTGTAAGTGTGAA 2111
414 ArgIleLeuGluAsnSerLeuLeuAsnLeuGluAsnGlyLysAlaGluIleValGlnLe 434
2112 TCMAAATGTTGAAGAAAGTGT-----GCTGAAGAAATTTGTAAGAAAGTGTAGCTGAA 2165
434 userAsp----- 436
2166 TGTGTAAGAAATTTGTAAGAAATTTGTAAGAAATTTGTAAGTGTGTAAGAAAT 2225
437 -PheSerLeuIleIleIleAspGluCysHisIleThrAsnLysGluAlaValIleTyrAsn 456
2226 TGTAGCTCAAGTGTGTGAAGAAAGTGTGCTCCAACTGTTGAAGAAAGTGTGAAAGAA 2285

456 nile-----MetArgIleIleMetGlnLysLeuLysAsnAs 469
2286 TGTGTAAGAAAGTGTAGCTGAAGAAATTTGTAAGAAATTTGTAAGTGTGAAAG 2345
469 ArgLeuLysLysGluAsnLysProValIlePro-----LeuProGlnIleLeuIleProVal 488
2346 TGTAGCTGAAGAAATTTGTAAGAAAGTGTAGCTCCAACTGTTGAAGAAATTTGTAAG 2405
488 AlaSerProGluValGlyValAlaThrLysGlnAlaLysAlaGluLysIleLeuLys 508
2406 TGTGTAAGAAAGTGTAGCTCCAACTGTTGAAGAAATTTGTAAGAAATTTGTAAG 2465
508 LysCysAlaAsnLeu----- 514
2466 TTTATGACCAATCTTTAAGTAATTTAATAGTGTATGCAAGCAAGCAAGCAAGCAAG 2525
514 Palaphe-----ThrIleLysThrValLysGluAsn-----LeuAspGlu 527
2526 CAGTATATTAAATGATAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2585
527 nLeuLysAsnGlnIleGlnGluProCysLysLysPheAla-----IleAlaAspAlaThrArg 546
2586 CGTAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2645
546 GGU-----AspProPheLysGluLysLeuLeuGluIle----- 557
2646 AGCAAAATCTATTACTAATGATCTATAGAGCAAAATTTGTAAGAAATTTGTAAGCAAG 2705
558 -----MetThrArgIleGlnThrLysCysGlnMetSerProMetSerAsp 573
2706 ATTAAGTGGCGCTTGAAGAAATTTGTAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2765
573 eGlyThrGlnProTyrGluGlnIleProAlaIle-----GlnMetGlu 586
2766 AATTGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2825
586 LysLysAlaAlaLysLys-----GlyAsnAlaGlyLysVal----- 599
2826 AGCAAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2885
600 -----CysAlaGlnIleLeuArgLysTyrAsnGluAlaLeu----- 611
2886 AAGTATGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2945
612 -----GlnIleAsnAspThrIleArgMetIleAspAlaIleThrHisLeuGlnIlePro 629
2946 ATTGATTAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 2990
629 eTyrAsnGluGluLysAspLysLysPheAlaValIleGluAspAspSerAspIleGlu 649
2991 AGCAAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 3050
649 yAspAspGluTyrCysAspGly----- 657
3051 ACAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 3110
657 pGluAspGluAspLeuLysLysProLeuLysLeuAspGluThrAspArgIleLeuMet 677
3111 ACAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 3129
677 eThrLeuPhePheGluAsnAsnLysMetLeuLysArgLeuAlaGluAsnProGluIleArg 697
3130 -----AATGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 3173
697 uAsnGlu-----LysLeuThrLysLeuArgAsnThrIleMetGluGlnIleThrAlaIleGlu 716
3174 GAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 3207
716 uGluSerAlaArgGlyIleIlePheThrLysThrArgGlnSerAlaIleValIleGlu 736
3208 -TCAAGTACTCAAGTGTTCAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 3246
736 nTrpIleThrLysGluAsnGluLysPheAlaGluValGlyLysAla-----HisIle 754

```

Db 3247 -----GTAGACAAAATGATATGTGATGTTGATCTCTGATACAAAGATCAAT 3299
QY 754 uilegly-----AlaglyHisSerSerGluPhe----- 763
Db 3300 TTATGGAATATTAAATGAGCAGGAGGAGGTTGAAAGAAATGTTTAAATTGCAAGATG 3359
QY 764 -----LysProMetThrC 768
Db 3360 ATTTAAAGTGAAGTGAATGTAATTACTGTGACAAAGATTAAGCATGTAATGGTTCAAAA 3419
QY 768 nasngluGlnylsGluValIleSerLysPheArgThrGlyLysIleAsnLeuLeuLeu 788
Db 3420 AGAGGTACAAAAAGAACTGTTAGTATTATGAAAGAAATGAAAGAAAT-----AlitC 3473
QY 788 aThrThrValAlaGluGluGlyLeuAspIleLysGluCysAsnIleValIleArgTyrC 808
Db 3474 AGATGATATTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3510
QY 808 yLeuValThrAsnGluIleAlaMetValGlnAlaArgGlyLysArgAlaArgAAspGluSe 828
Db 3511 -----AGATGATGATGCA-----CTAGAGGATG 3536
QY 828 rThrTyValLeuValAlaHisSerGlySerGlyValIleGluHisGluThrValAsnAs 848
Db 3537 CATAGCAATA-----TCTTCAGATCTTAACAGCAAGCAAGCAAGCAAGCAAGCAAGCA 3584
QY 848 pPheArgGluLysMetMetTyLysAlaIleHisCysValGlnAsnMetLysProGluC 868
Db 3585 T-----AAAGAAAAAGATGTTCACTAGTTCAGAACAGTTCAGCAATGATATGATGATGA 3641
QY 868 uTyAlaHisLysIleLeuGluLeuGluMetGlnSerIleMetGluLysLysMetLys 887
Db 3642 AAGTGTGAGAAAGTTTGAATTC-----AAAATATGAGAGAGAGCTAATGAGAGCA 3695
QY 887 ----- 887
Db 3696 TGCTGTGAATTAATGACATTACTAGCAAACTTATGAGAACTCAAGATTAATGA 3755
QY 888 -ThiLysArgAsnIleAlaLysHisTyLysAsnAsnProSerLeuIleThrPheLysC 907
Db 3756 AGTACAAACCAATTTAATTAAGATATGAGAAATTAAGAAATTAAGAAATTAAGAAATTA 3815
QY 907 sLysAsnGlySerValLeuAlaCysSerGlyGluAspIle-----HisValIleGlu 925
Db 3816 AGAATCTCTAAAGAAATTAATGATGCAAAAGATGATGATGATGATGATGATGATGATGAT 3875
QY 925 smetHisHisValAsnMetThr---ProGluPheLysGluLeuTyLysIleValArgGluAs 944
Db 3876 GGAACATCATTAACGACGAGCTGGATGAAAGTTGATGATTAAGAAATTAAGAAATTAAG 3935
QY 944 nLysAlaLeuGlnLysCysAlaAspTyLysIleAsnGluGluLeuIleCysLys 963
Db 3936 CAAGATC-----CAAAAAGTATGATTAAGATTCGAAAGATTAATTAAGAAATTA 1987

```

RESULT 15

```

US-08-973-462-1
; Sequence 1, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DROULHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973.462B
; EARLIER FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1

```

```

; LENGTH: 6152
; TYPE: DNA
; ORGANISM: P. falciparum
; US-08-973-462-1

Alignment Scores:
Score: 5.33e-10 Length: 6152
Percent Similarity: 186.50 Matches: 219
Best Local Similarity: 35.09% Conservative: 181
Query Match: 19.21% Mismatches: 390
DB: 3.51% Indels: 351
Gaps: 49

US-09-515-363c-2 (1-1025) x US-08-973-462-1 (1-6152)

QY 42 ValLysGluGlnIleGlnAlaArgThrValAlaThrSerGlyAsnMetGlnAlaValIleLeu 61
Db 1211 GTTGAAGAAAGTGTAGCTGAGAAATGTTGAAGAAAGTGTAGCTGAGAAATGTTGA 1270
QY 62 LeuLeuSerThrLeuGluGluGlyValIThrHisLeuLysThrArgGluProLysGlu 81
Db 1271 GTAGCTCCAACTGTTGAGAAATGCTA----- 1297
QY 82 AlaLeuArgArgThrGlySerProLeuAlaAlaArgTyrMetAsnProGluLeuThrAsp 101
Db 1298 -----GCTCCAACTGTTGAAGAAATGTTGAAGAAATGTTGAAGAAATGTTGA 1339
QY 102 LeuProSerProSerPheGluAsnAlaHisAspLysIleGluLeu----- 116
Db 1340 AGTGTGCTCCAACTGTTGAAGAAAGTGTAGAAAGAAATGTTGAAGAAAGTGTAGAA 1399
QY 116 ----- 116
Db 1400 AATGTTGAGAAAGTGTAGCTGAGAAATGTTGAAGAAAGTGTAGCTGAGAAATGTTGA 1459
QY 117 -----LeuLeuAsnLeuLeuPro 123
Db 1460 AGTGTAGCTGAATATGTTGAAGAAAGTGTAGCTGAATATGTTGAAGAAATGTTGA 1519
QY 124 ThrLeu-----ValAspLysLeuLeuValArgAspValIleAsp 136
Db 1520 ACTGTTGAGAAAGTGTAGCTGAGCTCAACTGTTGAAGAAATGTTGAAGAAATGTTGA 1579
QY 137 LysCysMetGluGluGluLeuLeuThrIleGluAspArgAsnArgIleAlaAlaGlu 156
Db 1580 AGTGTGCTCCAACTGTTGAAGAAAGTGTAGAAAGAAATGTTGAAGAAAGTGTAGAA 1639
QY 157 AsnAsnGlyAsnGluSerGlyValArgGluLeuLeuLysArgIleValGlnLysThrAsn 176
Db 1640 AAT-----GTTGAAGAAAGTGTAGCTGAGCTCAACTGTTGAAGAAATGTTGAAGAA 1693
QY 177 TrpPheSerAlaPheLeuAsnValLeuArgGlnThrLysAsnAsnGluLeuValIleGlu 196
Db 1694 GAAGAAAGTGTAGCTGAAGAAATGTTGAAGAAAGTGTAGCTGAAGAAAT-----GTTGA 1747
QY 197 LeuThrGlySerAspCysSerGluSerAsnAlaGluIleGluAsnLeuSerGlnLysAsp 216
Db 1748 AGTGTAGCTGAAGAAATGTTGAAGAAAGTGTAGCT-----GAAATGTTGAAGAAATG 1801
QY 217 GlyProGlnValGluGluGlnLeuLeuSerThr----- 227
Db 1802 GCTCCAACTGTTGAAGAAATGTTGAAGAAATGTTGAAGAAATGTTGAAGAAATGTTGA 1861
QY 228 -----ThrValGlnProAsnLeuGluLysGluValIThrPheMetGluAsnAsnSer 245
Db 1862 GTAGAAAGTGTGCTCCAACTGTTGAAGAAAGTGTAGCT-----GTTGAAGAAATGTTGA 1912
QY 246 GluSerSerPheAlaAspSerSerValValSerGluSerAspThrSerLeuAlaIleGlu 265
Db 1913 GAA-----AGTGTAGCTGAAGAAATGTTGAAGAAAGTGTAGCTGAAGAAAT 1954
QY 266 SerValSerCysLeuAspGluSerLeuGluIleHisAsnSerAsnMetGlySerAspSerGly 285

```

Db	1955	-----GTTGAGAAAGGTGAGCTGCAAAATCTT-----	198
OY	286	ThrMetGlySerAspSerAspGluAsnValAlaIaIaArgAlaSerProGluProGlu	305
Db	1982	-----GAAAGAAAGTGTAGCTGAGAAAGTTGAGAGAAATCTGACTCTCAACTGTGAA	2032
OY	306	LeuGluLeuArgProTyrGlnMetGluValAlaGlnProAlaLeuGluGlnGlyAsnIle	325
Db	2033	GAATTCGATAGCTCCAACTGTTGAGAAATTTGACTCCAAAGTGT-----CTA	2080
OY	336	IleIleCysLeuProThr-----GlySerGlyThrArgValAlaValTyrIleAla	343
Db	2081	GAAAGTGTGGCTCCAGAGTGTGAGAAAGAGTGTACAGAAAGAAATCTTCAACAAAGTGTAT	2144
OY	344	LysAspHisLeuAspLysLysLysLysAlaSerGluProGlyLysValIleValLeuVal	363
Db	2141	GAA-----AATGTTGAGAAAGAGTGTGAGAAATTTGTAAGAAATTTGATGCA	2191
OY	364	AsnLysValLeuLeuValGlu-GluLeuPheArgGlyGluPheGlnProIleLeuLysIly	383
Db	2192	AATGTTGAGAAAGAGTGTAGCTCCAACTGTGAGAAATCTGACTCCAAATGTTGAA	2251
OY	383	s-----TTPyrArgValIleGlyLeuSerGlyLysPth	394
Db	2252	ATTGTAGCTCCAACTGTTGTAGAAAGTGTGGCTCCAACTTTTGACAA-ACTGTACAA	2310
OY	394	GlnLeuLysIleSerPheProGluValValLysSerCysAspIleIleIleSerThrAl	414
Db	2311	AAATCTTGAGAAAGAGTGTACTGCAAAATGTTGAGAA-----AGTGTAC	2355
OY	414	AGlnIleLeuGluAsnSerLeuAsnLeuGluAsnGlnGlyGluAspAlaIleValGlnIle	434
Db	2356	TGAAATCTGTGAGAAAGTGTATA-----GCTGAAATGTGTGAGAAAGGTGTAGCTGAAA	2409
OY	434	userAsp-----	436
Db	2410	TGTTGAGAAAGATCTGACTCCAACTGTGAGAAATCTGACTCCAACTGTGAGAAAT	2468
OY	437	-PheSerLeuIleIleIleAspGluCysHisIleThrAsnLysGluAlaValTyrAsnAs	456
Db	2470	TGTAGCTCCAAAGTGTGTAGAAAGTGTGGCTCCAACTGTGAGAAAGGTGTAGAAAGAA	2529
OY	456	nIle-----MetArgIleTyrLeuMetGlnLysIleLysAsnAs	469
Db	2530	TGTTGAGAAAGGTGACTGCAAAATTTTGAGAAAGGTGTAGCTGAGAAATGTGAACAAAG	2589
OY	469	narGluLeuLysGluAsnLysProValIlePro-LeuProGlnIleGluGlyLeuTh	488
Db	2590	TGTAGCTCGAAATCTTGACAAAGAGTGTAGCTCCAACTGTGAGAAATTCAGCTCTAG	2649
OY	488	AlaSerProGlyValGlyValAlaThrLysGlnAlaLysAlaGluGlnIleIleLeuLys	508
Db	2650	TGTTGAGAAAGAGTGTACTCCAACTGTGAGAAAGTGTGTAAGAAAGTGTGTAAGAA	2709
OY	508	slenCysAlaAsnLeu-----As	514
Db	2710	TTTATCGACAAATCTTTAAGTATTTATTAGTGTATCGAATCGAAGCAAAATTAAGCA	2769
OY	514	palaPhe-----ThrIleSerThrValLysGluAsn-----LeuAspG	527
Db	2770	CAGTATATTAAATGAGATTGACAAAGATTAAGAAAGAAATGTAGTACCACCAATACTAGAAA	2829
OY	527	nleuLysAsnGlnIleGlnGluProCysLysLysPheAla-----IleAlaAspAlaThr	546
Db	2830	CCTGGAAGAAACTACAGCTGAAAGGTGTACTACTTTTACTTACATATTAGAGGATAC	2889
OY	546	gGlu-----AspProPheLysGluLysLeuLeuGluIle-----	557
Db	2890	AGAAATACTATTACTAATGATCTATTAGAGGAAAAATTAGAGAGACATCAGCAAAATGT	2949
OY	558	-----MetThrArgIleGlnIleThrTyrCysGlnMetSerProMetLysAspPh	573
Db	2950	ATTAGTGGCGCTTTGAAAAATATCCCAAAAGTGAACAGGAGAAAGAAAGAAATTAATGATGT	3009

OY	573	celythrlnprcyrgugntnrlpalalle-----gl-metG	586
Dd	3010	AATTGAAGACGACTAAAAGAAAGCGTCGTACCACCTTTAATAGAAGTCTGGACATAACGCA	3059
OY	586	uLysLysAlaAlaLysLys-----GLysAsnArgLysGluArgVal-----	599
Dd	3070	ACGAAAGAGGCCAAATACAAATTACGAAATATTTGAAATAATTTAGAAGAAATGGCATGGA	3129
OY	600	-----CysAlagUHisLeuArgLysTyraSnglValaleu-----	611
Dd	3130	AAAGTAATGAAATGCTGCAGAGCAATTTACGAAATTTAAACGAAACGTGTTTTAAATATCT	3189
OY	612	-----GlutlleasnsPThrIleArgMetIleaspAlaTyrrHrHisLeuGluThrPh	629
Dd	3190	ATTGATATTAACTAGACGAAACAGTAGCAAATTTACGGA-----GAAACITTT	3234
OY	629	eTryfAsnGluLulysAspLysPheAlaValIIegLusAspSerAsnGluTrgyl	649
Dd	3235	ACAAAACATCAATTAATGATATAGCACTTTTACTGMAATATTTTGATTAATGTAATGAAAT	3294
OY	649	YAspAspGluTrgycysAspGly-----As	657
Dd	3295	ACAACAAAAATTTATTACAGCTATGTTTCCAGATATAGAACCACTATATGTAATGTAATG	3354
OY	657	pGluAspGluAspAspleuLysLysProleuLysLeuAspGluThrAspArgThr-----	677
Dd	3355	ACGAACAAAGCGTTGATTTG-----	3373
OY	677	tThrleuPheheGluLusnsnLysMetLeuLysArgLeuAlaGluAsnProGluTrgyl	697
Dd	3374	-----AATGAAATGGGGTTACTTCGATTTAGATTAATATAGAACATAT	3417
OY	697	uAsnGlu----LysLeuthrHrLysLeuArgAsnThrIleMetGluGlnTyrrHrArIleHgl	716
Dd	3418	GAAACAGCTTATTAAATATATAGAAAAATAT-----	3451
OY	716	uCluSerAlaArgGlyIleIlePheHrLysThrArgGlnSerAlaTyrrAlaIleuHergl	736
Dd	3452	TCAAGTACTGCAAGCTGTTCAAGAAACGTAACTGACAT-----	3490
OY	736	nTrpIleThrcGluAsnGluLysPheAlaGluValIGlyValLysAla-----HisHisLe	754
Dd	3491	-----GTAAACAAAAATGATATATATGCGATGTGCTGCTGCTATGAAGAAATAT	3543
OY	754	uIleGly-----AlaGlyHisSerSerGluPhe-----	763
Dd	3544	TTTAGCAATATTAAATGAGCGACGAGCGTTGAAACAAATGTTTTTAATTTGCAATATG	3603
OY	764	-----LysPromEIIIHgl	768
Dd	3604	ATTTAAAGTCGAAGTGATTAATTTACTGTACAGCAAAATTAAGATGTGAACCGGT	3663
OY	768	nasnGluGluLysGluValIIeserLysPheArgThrGlyLysIleasnLeuAlleal	788
Dd	3664	AGAGGTGAAAAAAAGAACTGTAGCTATTATTCAGAAATGAGAACAAAT-----ATTGT	3717
OY	788	aThrThrValAlaGluGluLysLeuAspLysGluGlySasnlIleValIIeal-tyrGI	808
Dd	3718	AGATGTATTAGGACAAACAAAACAAATTTTAAACAC-----	3754
OY	808	yLeuValIThrAsnGluIleAlaMetValIlnlaIleArgGlyArgAlaArgAlaAspGluSe	828
Dd	3755	-----AAGTGATTAATATCA-----GTACAAATATC	3780
OY	828	rThrTyrrValLeuValAlaHisSerGlySerGlyValIIlegUHisGluThrValAsnAs	848
Dd	3781	CATGCAATAT-----TCTTCAGATCTTAAAGAGAAACCTGATCATATTAAGA	3828
OY	848	pHeArgGluLysMetMetTyrrLysAlaIIehisGlyValGlnsnMetLysProGluG	868
Dd	3829	T---AAAGAAAGATGCTTTCACCTACTGCTTGAACAAGTATGAACATGATATATGCA	3885

```
QY 868 uTyrAlaHisLysIleLeuGluMetGlnSerIleMetGluLysIleMetLys-- 887
      |||:|||||
      |||:|||||
Db 3886 AAGTGTGAGAAAGTTTACAAATG-----AAAAATATGCAAGAGAGGATTAATGAGCA 3939
      |||:|||||
      |||:|||||
QY 887 ----- 887
Db 3940 TGCCTTGAAATTAATGACATTACTAGCAAACTTATGAGAAACTCAGAACTTAATGA 3999
QY 888 -ThiLysArgAsnIleAlaLysHisTyrIleLysAsnAsnProSerLeuIleIlePheLeuCy 907
      |||:|||||
      |||:|||||
Db 4000 AGTAGAAGCAGATTTAATTAAGCATATGCAAAATTAAGAAATTGAGAAAGCATTATC 4059
      |||:|||||
      |||:|||||
QY 907 sLysAsnCySerValLeuAlaCySerGlyIleAspIle-----HisValIleGluLys 925
      |||:|||||
      |||:|||||
Db 4060 AGAAGATTCTAAGAAATTAATATAGATGCAAAAGATGATACATTAGAAAAGCTTATGAGCA 4119
      |||:|||||
      |||:|||||
QY 925 smethHisValAsnMetThr---ProGluPheLysGluLeuTyrIleValArgGluAs 944
      |||:|||||
      |||:|||||
Db 4120 GGAACATGATATACGACGAGCTGGATGAAGTTGAAATTAAGATGCGAAGAGAGA 4179
      |||:|||||
      |||:|||||
QY 944 nLysAlaLeuGluLysLysCysAlaAspTyrGlnIleAsnGlyGluIleIleCysLys 963
      |||:|||||
      |||:|||||
Db 4180 CAAAGATC-----GAAAAAGTATCTGATTTAAAGATCTTGAAAGAGATA ATTAAA 4231
      |||:|||||
      |||:|||||
```

Search completed: May 16, 2003, 16:35:43
Job time : 328 secs

GenCore version 5.1.4 p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 16, 2003, 10:02:59 ; Search time 593 seconds

(without alignments)
6539.089 Million cell updates/sec

Title: US-09-515-363c-2

Perfect score: 5311
Sequence: 1 MSNGXSTDENRYLISCRA.....LPITFPNDYSTVCLFSD 1025

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Command line parameters:

-MODEL-frame_p2n.model -DEV-x1h
-O/cgn2.1/USPTO_SPOOL/US09515363/runat.16052003.100245.24779/app_query.fasta.1.1223
-DB-N.Geneseq.101002 -OFMT-fastlap -SUFFIX-frag -MINMATCH-0.1 -LOOPEXT-0
-LOOPEXT-0 -UNITS-bits -START-1 -END-1 -MATRIX-Biosum62 -TRANS-human:40 -cdl
-LIST-45 -DOCALIGN-200 -THR SCORE-pct -THR MAX-1008 -THR MIN-0 -ALIGN-15
-MODE-LOCAL -OUTFMT-prot -NORR-ext -HEAPSIZE-500 -MINLEN-0 -MAXLEN-200000000
-USER-US09515363.ecgn1.1.1229 -runat.16052003.100245.24779 -NCPU-6 -ICPU-3
-NO_XLPHY -NO_MAP -LARGEOUERY -NEG_SCORES-0 -WA-T -LONLIDG -DEV_TIMEROUT-120
-NARN_TIMEROUT-30 -THREADS-1 -XGAPOP-10 -XGAPEXT-0.5 -Fgapop-6 -Fgapext-7
-YGAPOP-10 -YGAPEXT-0.5 -DELOP-6 -DELEXT-7

Database :

N.Geneseq.101002: *
1: /SID2/gcgcdata/geneseq/geneseqn-emb1/NA1980.DAT: *
2: /SID2/gcgcdata/geneseq/geneseqn-emb1/NA1981.DAT: *
3: /SID2/gcgcdata/geneseq/geneseqn-emb1/NA1982.DAT: *
4: /SID2/gcgcdata/geneseq/geneseqn-emb1/NA1983.DAT: *
5: /SID2/gcgcdata/geneseq/geneseqn-emb1/NA1984.DAT: *
6: /SID2/gcgcdata/geneseq/geneseqn-emb1/NA1985.DAT: *
7: /SID2/gcgcdata/geneseq/geneseqn-emb1/NA1986.DAT: *
8: /SID2/gcgcdata/geneseq/geneseqn-emb1/NA1987.DAT: *
9: /SID2/gcgcdata/geneseq/geneseqn-emb1/NA1988.DAT: *
10: /SID2/gcgcdata/geneseq/geneseqn-emb1/NA1989.DAT: *
11: /SID2/gcgcdata/geneseq/geneseqn-emb1/NA1990.DAT: *
12: /SID2/gcgcdata/geneseq/geneseqn-emb1/NA1991.DAT: *
13: /SID2/gcgcdata/geneseq/geneseqn-emb1/NA1992.DAT: *
14: /SID2/gcgcdata/geneseq/geneseqn-emb1/NA1993.DAT: *
15: /SID2/gcgcdata/geneseq/geneseqn-emb1/NA1994.DAT: *
16: /SID2/gcgcdata/geneseq/geneseqn-emb1/NA1995.DAT: *
17: /SID2/gcgcdata/geneseq/geneseqn-emb1/NA1996.DAT: *
18: /SID2/gcgcdata/geneseq/geneseqn-emb1/NA1997.DAT: *
19: /SID2/gcgcdata/geneseq/geneseqn-emb1/NA1998.DAT: *
20: /SID2/gcgcdata/geneseq/geneseqn-emb1/NA1999.DAT: *
21: /SID2/gcgcdata/geneseq/geneseqn-emb1/NA2000.DAT: *
22: /SID2/gcgcdata/geneseq/geneseqn-emb1/NA2001.DAT: *
23: /SID2/gcgcdata/geneseq/geneseqn-emb1/NA2002.DAT: *
24: /SID2/gcgcdata/geneseq/geneseqn-emb1/NA2002.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5311	100.0	3365	AA017203	Human RNA helicase
2	5285	99.5	3372	ABA04908	Human RNA helicase
3	5276.5	99.4	3131	AA011170	Human melanoma dif
4	3176	59.8	1967	AA040960	CDNA encoding nove
5	2442	46.0	1443	ABA04916	Human RNA helicase
6	2356.5	44.4	1557	AA159285	Human polynucleoti
7	2120	39.9	1382	AA161071	Human polynucleoti
8	1995	37.6	1284	ABA04913	Human RNA helicase
9	1961.5	36.5	1319	AA041517	CDNA encoding nove
10	1344	25.3	2613	AA015201	Human cDNA sequenc
11	1201.5	22.6	1258	AA001149	Interferon induced
12	1201.5	22.6	1270	AA001151	Interferon induced
13	1106.5	20.8	956	AA091688	CDNA encoding nove
14	991	18.7	1978	AA040969	CDNA encoding nove
15	968.5	18.3	3692	AA001150	Interferon induced
16	969.5	18.3	3704	AA001152	Interferon induced
17	830.5	15.6	1346	AA041524	CDNA encoding nove
18	783	14.7	499	AA022046	Human foetal liver
19	783	14.7	499	AA022046	Human brain expres
20	783	14.7	499	AA022046	Human bone marrow
21	783	14.7	499	AA022046	Human bone marrow
22	783	14.7	499	AA022046	Human bone marrow
23	783	14.7	499	AA022046	Human bone marrow
24	783	14.7	499	AA022046	Human bone marrow
25	783	14.7	499	AA022046	Human bone marrow
26	782	14.7	609	AA033745	Human colon cancer
27	782	14.7	609	AA033745	Human ovarian anti
28	708	13.3	3026	AA018139	Human cDNA sequenc
29	666.5	12.5	1600	AA017190	Human ORF2745
30	666	12.5	1036	AA017204	Human melanoma dif
31	666	12.5	6406	AA017207	Human melanoma dif
32	649	12.2	374	AA069625	EST clone COS22.
33	642	11.9	392	AA069625	Novel human polyu
34	630	11.9	387	AA069625	Novel human polyu
35	563.5	10.6	3184	ABA04248	Human ovarian anti
36	553	10.4	340	AA14761	Human breast cance
37	553	10.4	341	AA123623	Human breast cance
38	498	9.4	458	AA047057	Nucleotide sequenc
39	488	9.2	292	ABA07057	Human breast cell
40	488	9.2	292	ABA04940	Human foetal liver
41	488	9.2	292	ABA04940	Probe #10515 for g
42	488	9.2	292	AA032049	Human brain expres
43	488	9.2	292	AA032049	Human bone marrow
44	488	9.2	292	AA032049	Probe #9846 for qe
45	488	9.2	292	AA032049	Probe #13794 used

ALIGNMENTS

RESULT 1

ID AAD17203 standard; CDNA: 3365 BP.

AC AAD17203:

DT 29-NOV-2001 (first entry)

XX Human melanoma differentiation associated (Mda)-5 cDNA.

DE Human: melanoma differentiation associated gene; Mda-5; interferon; IFN;

KW RNA helicase motif; caspase recruitment domain; CARD; therapy; melanoma;

KW neuroblastoma; astrocytoma; glioblastoma; multifactor; cancer; cervical;

KW breast; colon; prostate; osteosarcoma; chondrosarcoma; systemic toxicity;

OS central nervous system; cytosolic; apoptosis; ss.

XX Homo sapiens.

XX

```

FH Key Location/Qualifiers
FT CDS 169..3246
FT     /tag- a
FT     /product= "Human melanoma differentiation associated
FT     (Mda)-5 protein"
FT misc_feature 3225..3228
FT     /tag- b
FT     /note= "ATTTA motif"
FT misc_feature 3284..3287
FT     /tag- c
FT     /note= "ATTTA motif"
FT polyA_signal 3343..3348
FT     /*tag- d
XX WO200164707-A1.
XX PD 07-SEP-2001.
XX PF 28-FEB-2001; 2001WO-US06960.
XX PR 29-FEB-2000; 2000US-0515363.
XX PA (UYCO ) UNIV COLUMBIA NEW YORK.
XX PI Fisher PB, Kang D, Gopalakrishnan RV;
XX WP1: 2001-565494/63.
XX DR P-PDB: MAE10155.
XX PT Nucleic acid sequences encoding a Melanoma Differentiation Associated
XX anti-viral activity -
XX PS Claim 1; Page 16-18; 152pp; English.

The present invention relates to an isolated nucleic acid encoding a
melanoma differentiation associated gene-5 (Mda-5) polypeptide. Mda-5
contains a caspase recruitment domain (CARD) and a RNA helicase motif.
Mda-5 is a novel interferon (IFN) inducible gene with structural
similarities to RNA helicases and CARD motifs containing proteins. Mda-5
is induced during terminal differentiation in human melanoma cells
treated with the combination of recombinant fibroblast IFN and the
anti-leukemic compound mezerein (MEZ). Mda-5 is useful for identifying
compounds that may induce its expression. Mda-5 is useful for treating
cancer e.g., melanoma, neuroblastoma, astrocytoma, glioblastoma
multiforme, cervical cancer, breast cancer, colon cancer, prostate
cancer, osteosarcoma, chondrosarcoma and especially where the cancer is
a cancer of the central nervous system and apoptosis. The Mda-5 promoter
exhibits melanocyte tissue specificity and minimises systemic toxicity.
The present sequence is human Mda-5 cDNA.

SO Sequence 3365 BP; 1138 A; 644 C; 753 G; 830 T; 0 other;

Alignment Scores:
Pred. No.: 0 Length: 3365
Score: 5311.00 Matches: 1025
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0

US-09-515-363C-2 (1-1025) x ADI17203 (1-3365)
QY 1 MetSerAnGlyTYrSerThrASPGLVAsnPhenArgTyrLeuLleSeRcySPheARdAla 20
Db 169 ATGTGCATGGGTATTCACAGACAGAGAATTCGCCGATATCTACTCGTATTTCAAGGCC 228
QY 21 ArgValLySMeTyrTrlIegInVaIGluProValLeuAspTyrlLeuthrPheLeuProAla 40
Db 229 AGCGTAAATAATGTACATCCAGGTGACGCTGTGCTGAGACTACTGACATTTCTGCTGCA 288
QY 41 GIUALLYSGUGLnllleclnArGrThrValAlahrserrGIyaaNmMeIShAlaValGln 60
|||||

```

Dh	289	GAGGTGAAGGACGATTTCAGACAGCTGCCACCTCCGGGACACATGACGCATCTTGA	348
Oy	61	LeuLeuLeuSerThrLeuGluLysGlyValTPrhISLeuGlyTPrhTArgGluIleVal	80
Dh	349	CTGCTGCTGACACCTCTGGACGAAGCGAGCTGGCACCTTGTTGGACTCGCAATTCGTG	408
Oy	81	GluAlaIleuArgThrGlySerProLeuAlaAlaArgTyrMetAspProGluLeuThr	100
Dh	409	GAGGCCCTCCGGAAACCGGACACCTCTGGCCCGCCGTACATCAACCTCAGCTACG	468
Oy	101	AspLeuProSerProSerPheGluAsnAlaHisAspGluTyrLeuGluIleuLeuAsnLeu	120
Dh	469	GACTGGCCCTCTGCATCTGTTTGAGAACCGTCATGATGAATATCTCCAACTGTGACCTC	528
Oy	121	LeuGluProThrLeuValAspLysLeuLeuValArgAspValIleuAspLysTyrMetGlu	140
Dh	529	CTTCAGCCCACTCGGTGGCAACGCTTAGTTAGAGACTCTTGATAGTGCATGCATGAG	588
Oy	141	GluGluLeuLeuThrIleGluAspArgAsnArgIleAlaAlaAlaGluAsnAsnLysAsn	160
Dh	589	GAGAACCTGTTGCACAAATTTGAAGACACAAACCGGATTCGTCGACGAAACAAATGGAAAT	648
Oy	161	GluSerGlyValArgGluLeuLeuLysArgIleValGluLysGluAsnTrpPheSerAla	180
Dh	649	GAACTACAGGTGTAAGAGAGCTACTAAAAAGAGATTGTCCACAAAGAAATGCTTCCTGCA	708
Oy	181	PheLeuAsnValLeuArgGlnThrGlyAsnAsnGluLeuValGlnGluLeuThrLysSer	200
Dh	709	TTTCTGAATGTTCTTGCTCAACAGGAAACAAATGAACCTGTCCAGAGATTACAGCTCT	768
Oy	201	AspLysSerGluSerAsnAlaGluIleGluAsnLeuSerGlnValAspGlyTyrThrVal	220
Dh	769	GATTGCTCAGAAACCAATGCGAGATTGGAATTTATCCAAAGTTGATGCTTCAATG	828
Oy	221	GluGluGlnLeuLeuSerThrThrValGlnProAsnLeuGluLysGluValTPrhMet	240
Dh	829	GAAAGCAACTCTTTTCAACACAGCTTCACCCAAATCTGGACAGAGCTCTGCTCATG	888
Oy	241	GluAsnAsnSerSerGluSerSerPheAlaAspSerSerValValSerGluSerAspThr	260
Dh	889	GAGAAATACATCATCAGCATATCTTTTGGCAGATTCTCTGTTGATTCCAGATCAACCA	948
Oy	261	SerLeuAlaGluGlySerValSerCysLeuAspGluSerLeuGlyHisAsnGlnMet	280
Dh	949	AGTTTGCCACAAACCAACTGTCACCTCTTAGATGAAGACTTGGACATACACCTGATG	1008
Oy	281	GlySerAspSerGlyThrMetGlySerAspSerAspGluGluAsnValAlaLeuArgAla	300
Dh	1009	GGCAGTATTCAGGCACCACTGAGGGAAGTGAATGTCAGATGAAGAAATGTGCAATACCA	1068
Oy	301	SerProGluProGluLeuGlnLeuArgProTyrGlnMetGluValAlaGluIleValLeu	320
Dh	1069	TCCCGGAGACAGACACTCCAGCTCAGGCGCTTACCAATGGAAGTTGCCAGCTAGCTTG	1128
Oy	321	GluGlyLysAsnIleIleIleCysLeuProThrGlySerGlyLysThrArgValAlaVal	340
Dh	1129	GAAAGGAAGATATCATCACTCTCCCTTACAGGAGTGGAAAAACAGATCTCTGTT	1188
Oy	341	TyrIleIleLysAspHisLeuAspLysLysLysLysAlaSerGluProGlyLysValIle	360
Dh	1189	TACATTGCCAAGATCTACTTAGACAAAGAAAAAGCACTTGACCTCTGGAAAAATATA	1248
Oy	361	ValLeuValAsnLysValLeuLeuValGluGlnLeuPheArgLysGluPheLeuProPhe	380
Dh	1249	GTTCTGTCATATAAGTACGTACGTAGTTGAACAGCTCTTCGGACAGAGTTCACACTTT	1308
Oy	381	LeuLysLysTrpTyrArgValIleGlyLeuSerGlyAspThrGlnLeuLysLysSerPhe	400
Dh	1309	TTGAAAGAAATGATCTGTTATTTAGATTAAAGGTGATACCAACTGAATAATCATATT	1368
Oy	401	ProGluValValLysSerCysAspIleIleIleSerThrAlaGlnIleLeuIleAsnSer	420
Dh	1369	CCAGAACTGTGTCAAGTCTCTGTATTATTATTCAGTACAGCTCAAAATCCTTGAAGATCC	1428

FT /product= "Human RH16"

XX MO200185955-A1.

XX 15-NOV-2001.

XX 11-MAY-2001: 2001WO-FR01441.

XX 11-MAY-2000: 2000FR-0006030.

XX (ISTA-) ISTAC.

XX (INSTP) INST PASTEUR LILLE.

XX Bahr G, Cocude C, Capron A;

XX WPI: 2002-082898/11.

XX P-PSDB: AAM47798.

XX New polypeptide, useful for treating and diagnosing cancer or

PT inflammation, and drug screening, comprises a human polynucleotide

XX homologous to RMA helicase.

XX Claim 7: Page 85-89; 114pp; French.

XX

CC The present sequence is the coding sequence for human RH16. RH16 is a

CC 116kDa protein and has homology to RMA helicases (DEXH box). RH16 and

CC its coding sequence are useful for treating cancer; acute or chronic

CC infections (especially by HIV or hepatitis B or C); inherited genetic

CC diseases (auto)immune diseases (particularly rheumatism, arthritis,

CC arteriosclerosis, osteoporosis and diabetes, but many others listed) and

CC to prevent graft rejection. RH16 and its coding sequence are also useful

CC for inducing, or increasing, the immune response to a vaccine.

XX

SQ Sequence 3372 BP: 1157 A; 637 C; 748 G; 830 T; 0 other:

Alignment Scores:

Pred. No.: 0 Length: 3372

Score: 5285.00 Matches: 1020

Percent Similarity: 99.71% Conservative: 2

Best Local Similarity: 99.51% Mismatches: 3

Query Match: 99.51% Indels: 0

DB: 24 Gaps: 0

US-09-515-363c-2 (1-1025) x ABA04908 (1-3372)

QY 1 MetSerAsnGlyTyrSerThrAspGluAsnPheArgTyrLeuIleSerCysPheArgAla 20

DB 155 ATGTCGAAATGGGATATCCACAGACAGCAATTCCTGATCTCATCTCGTTCGAGGACC 214

QY 21 ArgValLysMetTyrIleGlnValGluProValLeuAspTyrLeuThrPheLeuProAla 40

DB 215 AAGGTGAATAATGATACATCCAGGTGGAGCTGTGCTGCTGACTACCTTCTGCTGCA 274

QY 41 GluValLysGlnIleGlnArgThrValAlaThrSerGlyAsnMetGlnAlaValGlu 60

DB 275 GAGGTGAAGACAGATTCACAGACAGACAGTCCGACCTCCGGGAACATGACAGCAGTTGA 334

QY 61 LeuLeuLeuSerThrLeuGlnLysGlyValTyrPheLysLeuGlyTyrThrArgLysPheVal 80

DB 335 CTGCTCTGACGACCTTGAGAGAGGAGTGTGACCTGTTGACCTGCGCAATTCCTG 394

QY 81 GluAlaLeuArgArgThrLysSerProLeuAlaAlaArgTyrMetAsnProGluLeuThr 100

DB 395 GAGGCGCTCCGAGAACCCGACGCTCTGGCCGCCGCTACATCAACCTCGAGCTCAGC 454

QY 101 AspLeuProSerProSerPheGluAsnAlaHisAspGlyTyrLeuGlnLeuLeuAsnLeu 120

DB 455 GACTTCGCTCTGTCATCGTTTGAAGAGCTCATGATGATATCTCCAACTGCTGAACTCC 514

QY 121 LeuGlnProThrLeuValAspLysLeuLeuValArgAspValLeuAspLysCysMetGlu 140

DB 515 CTTCAACCCACATCTGTGTGACAAAGCTTCTTAGAGAGAGCTTGTGATACGTCATGGAG 574

QY 141 GluGluLeuLeuThrIleGluAspArgAsnArgIleAlaAlaGluAsnGlyAsn 160

DB 575 GAGCAACTGTTCACATATGAGACAGAAACCGAGATTGCTGCGCAAAACAAATGAAAT 634

QY 161 GluSerGlyValArgGluLeuLeuLysArgIleValGlnLysGluAsnThrPheSerAla 180

DB 635 GATACAGGTGTAGAGAGCTACTATAAAGATGTGGCAAGAAAGAAACTGGTGTGTGCA 694

QY 181 PheLeuAsnValLeuArgGlnThrGlyAsnAsnGluLeuValGlnGluLeuThrIleSer 200

DB 695 TTTCGAATGTCTCTGTCGCAAAACAGAAACATGAACTGTGCAAGATTAACAGCTCT 754

QY 201 AspCysSerGluSerAsnAlaGluIleGluAsnLeuSerGlnValAspGlyProGluVal 220

DB 755 GATTGCTCAGAAAGCAATGCAGAGATTGAGAAATTATTCACAAAGTTGATGTCTCTAAGTG 814

QY 221 GluGlnGlnLeuLeuSerThrValGlnProAsnLeuGlnLysGluValTyrGlyMet 240

DB 815 CAGAGCAACTTCTTTCACACACAGCTTACGCCAAATCTGGAGAGAGAGCTGTGAACTG 874

QY 241 GluAsnAsnSerSerGluSerSerPheAlaAspSerSerValSerGluSerAspThr 260

DB 875 GAGAAATACCTCATGATCATCTTTCGACATTTCTGTAGTTTCAGATTCATACACACA 934

QY 261 SerLeuAlaGluGlySerValSerCysLeuAspGluSerLeuGlyHisAsnMet 280

DB 935 AATTTGGCAGAAAGCAAGTGCAGCTGCTTATGATGAACTGTGGACATTAACATTAACATG 994

QY 281 GlySerAspSerGlyThrMetGlySerAspSerAspGluGluAsnValAlaValArgAla 300

DB 995 GGCATGTATTACAGCACCATGGAAAGTATTCAGATGAAGAAATGTGGCAATTAACACA 1054

QY 301 SerProGluProGluLeuGlnLeuArgProTyrGlnMetGluValAlaGlnThrAlaLeu 320

DB 1055 TCCCCGAGACGAGAACTCCAGCTCAGGCTTACCAATGGAAGTGGCCACATTAACCTTG 1114

QY 321 GluGlyLysAsnIleIleIleCysLeuProThrGlySerGlyLysThrArgValAlaVal 340

DB 1115 GAAAGGAAAGAAATATCATATCTGCTCCCTCAGAGGAGTGGAAAAACCGAGATCTGCTT 1174

QY 341 TyrIleAlaLysAspHisIleAspLysLysLysLysAlaSerGluProGlyLysValIle 360

DB 1175 TACATTGCCCAAGATCATCTTACACAGAAAGAAAAAGCATCTGAGCCCTGGAAGAAATTATA 1234

QY 361 ValLeuValAsnLysValLeuLeuValGlnGlnLeuPheArgLysGluPheGlnProPhe 380

DB 1235 GTTCTTGCAATTAAGCTACTGCTACTGTAACAGCTCTCCGCAAGAGATTCACATTT 1294

QY 381 LeuLysLysTyrTyrArgValIleGlyLeuSerGlyAspThrGlnLeuLysIleSerPhe 400

DB 1295 TTGAAGAAATGATCGTGTATTGATTAAGTGTGTATACCCCACTGAATAATATCATTT 1354

QY 401 ProGluValValLysSerCysAspIleIleIleSerThrAlaGlnIleLeuGluAsnSer 420

DB 1355 CCAGAAAGTTCAGAGCTCTGTGATTTTATTCAGTACAGCTCAAACTCTGAAAGCTCC 1414

QY 421 LeuLeuAsnLeuGluAsnGlyLysAspAlaGlyValGlnLeuSerAspPheSerIle 440

DB 1415 CTCTTAAACTTGGAAAAAGGAGAGATCTGCTGCTCAATGTAGACTTTCCTCATTT 1474

QY 441 IleIleAspGluCysHisIleThrAsnLysGlnAlaValTyrAsnAsnIleMetArgHis 460

DB 1475 ATCATTTGATGATTCATCATCACCAACCAACACAGTGAATTAATTCATCATCAAGCAT 1534

QY 461 TyrLeuMetGlnLysLeuLysAsnAsnArgLeuLysLysGluAsnLysProValIlePro 480

DB 1535 TATTGATGCGAAGATTAAAAACATAGACTCAAGAAAGAAACCAACCATGATTTCC 1594

QY 481 LeuProGlnIleLeuGlyLeuThrAlaSerProGlyValGlyGlyAlaThrLysGlnAla 500

DB 1595 CTCTCTAGATAGTGGAGCTAAACAGCTTCACTGCTGTGTGAGGAGGCGCAGCAAGAGCC 1654

QY 501 LysAlaGluGlnHisIleLeuLysLeuCysAlaAsnLeuAspAlaPheThrIleLysThr 520

Db 1678 ATTCAACTTATTGCAATGAGTCCCATGTCAGATTTCGAAATCCAAATATATGAAACA 1737
 QY 581 TTPAlaIleGluMetGluLysAlaAlaLysGlySerThrGlySerThrValLys 600
 Db 1738 TGGGCCATTCAAATGAAAAAAGCTGCAAAAAAGGAATGCAAGACAGCTGTTC 1797
 QY 601 AlAGluHisLeuArgLysTyrAsnGluAlaLeuGlnIleAsnSerPheThrLeuArgMetGlu 620
 Db 1798 GCGAACAATTGAGAGAGACATGAGCCCTACAAATTAATACACCAATCCAAATGATA 1857
 QY 621 Asp-AlaTyrThrHisLeuGluThrPheTyrAsnGluLysAspLysPheAlaVal 640
 Db 1858 GATCCCGATTAATCTCATCTTGAAGAACTTCTATTAATGACAGAAAGATAAAGATTG 1917
 QY 640 AlIleLeuAspSerSerAspGluGlyLysAspSerLysGlySerAspGluAspG 660
 Db 1918 TCATTAAGATGATGATGATGAGGGGATGATGATGATGATGATGATGATGATGATG 1977
 QY 660 LuAspAspLeuLysLysProLeuLysLeuAspGluThrAspAlaGlyPheLeuP 680
 Db 1978 AGGATATTTAAAGAAACCTTTGAAACCTGATGAAACACATTAATTCATGACTTAT 2037
 QY 680 hepGluAsnAsnLysMetLeuLysArgLeuAlaGluAsnProGluLysGluLys 700
 Db 2038 TTTTGAAAAACATTAATGTTGAAAAAGCTGCTGAAAGCCAGAAATAGAAAAAGAA 2097
 QY 700 yslLeuThrLysLeuArgAsnThrIleMetGluGlnTyrThrAlaGlnLysLeuAla 720
 Db 2098 AGCTGACCAATTAAGAAATACCATATAGAGAAATACATGAACTGAGAAATGACAC 2157
 QY 720 rglLysIleIlePheThrLysThrArgLysSerAlaTyrAlaLysSerGlnTyrIle 740
 Db 2158 GAGGATTAATCTTACAAAAACGACAGACATGATGCTGCTTCCCATGATTAATG 2217
 QY 740 LuAsnGluLysPheAlaGluAlaGlyValLysAlaHisLeuLeuGlyAlaGlyHis 760
 Db 2218 AAAATGAAAAATTTGCTGAAGTAGAGATCAAGCCCATCATGATGAGGCTGAGACACA 2277
 QY 760 erSerGluPheLysProMetThrGluAsnGluLysGluValIleSerLysPheArg 780
 Db 2278 GCAGTAGTTCAAAACCATCAGACAGAAATGAAACAAAGACATTAATTAATTTGCA 2337
 QY 780 hrcLysLysIleAsnLeuLeuIleAlaThrValAlaGluGluGlyLeuAspIleLys 800
 Db 2338 CTGCAAAAATCATCTCTGCTTATGCTGCTACCAATGCAAGAGAGAGATTAATTAAG 2397
 QY 800 LuCysAsnIleValIleArgTyrGlyLeuValThrAsnGluIleAlaMetValGluAla 820
 Db 2398 AATGTAAACATTTTATCCGTTATGCTGCTGCTACCAATTAATTAAGCATGCTCAGCC 2457
 QY 820 rglLysArgAlaArgAlaAspGluSerThrTyrValIleValAlaHisSerGlySer 840
 Db 2458 GTGGTGCAGCAGAGCTGATGAGACAGACCTACCTCTGCTGCTGCTGCTGCTGCTG 2517
 QY 840 AlIleGluHisGluThrValAsnAspPheArgGluLysMetGlyTyrAlaIleHis 860
 Db 2518 TTATCGAACATGACAGCATTAATGATTTCCGACAGACATGATTAATTAAGCTTACAT 2577
 QY 860 yslValGluAsnMetLysProGluLysAlaHisLysLeuLeuGluLysGluMetCys 880
 Db 2578 GTGTTCAAATATGAAACCGACAGAGATGCTCATTAAGATTTGAAATTCAGATGCAAA 2637
 QY 880 erIleMetGluLysLysMetLysThrLysArgAsnIleAlaLysHisValLysAsnAsp 900
 Db 2638 GTATTAATGAAAAAAGAAATTAACCAAGAAATATTCGCAAGCAATTAAGAAATTAAC 2697
 QY 900 roserLeuIleThrPheLeuCysLysAsnGlySerValIleAlaCysSerGlyLysAsp 920
 Db 2698 CATCACTAATTAATCTTCTTTCGAAAAACCTGAGTGTGCTGCTGCTGCTGCTGCTG 2757
 QY 920 leuHisValIleGluLysMetHisValAsnMetThrProLysPheLysGluLeuTyr 940
 Db 2758 TCATGTAAATGAGAAAAATGATCAGCTCAATTAATGACCCCAATATTCATCAATTAAT 2817

QY 940 leuValArgGluAsnLysAlaLeuGlnLysCysAlaAspTyrGlnIleAsnGlu 960
 Db 2818 TTGTAAAGAAACAAAGACATGCAAAAGAAAGTGTGCGACTATTAATTAATGATGAA 2877
 QY 960 leIleCysLysCysGlyGlnAlaTyrGlyThrMetMetValHisLysGlyLeuAspLeu 980
 Db 2878 TCATCTGCAATGATGCGCCAGCTTGGGACACATATGATGTCACAAAGCTTACATTTGC 2937
 QY 980 rocLysLeuLysIleArgAsnPheValValPheLysAsnAsnSerThrLysGlnT 1000
 Db 2938 CTGTCTCAAAATTAAGAAATTTGTAGTGTGTTTCAAAAATTAATTCACAAAGAA 2997
 QY 1000 yrlLysLysTyrPalGluLeuProIleThrPheProAsnLeuAspTyrSerGly 1020
 Db 2998 ACAAAAAGGGGTAGAAATTAATCCATTCACATTTCCCAATCTTCACTATTACAAAT 3057
 QY 1020 eupHeserAspLysAsp 1025
 Db 3058 TATTAGTATGAGGAT 3074
 Db
 RESULT 4
 AAS40960
 ID AAS40960 standard; cDNA: 1967 BP.
 XX
 AC AAS40960;
 DT 17-DEC-2001 (first entry)
 DE
 CDNA encoding novel human enzyme polypeptide #176.
 KW Human: oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
 KW ligase; hyperproliferative disorder; immunodeficiency disorder;
 KW autoimmune disorder; neurological disorder; metabolic disorder;
 KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
 KW blood-related disorder; infectious disorder; gene therapy; cytostatic;
 KW anti arthritic; nephrotropic; anticoagulant; ss.
 OS Homo sapiens.
 PN W020015301-A2.
 PD 02-AUG-2001.
 PE 17-JAN-2001; 2001WO-US01239.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000: 2000US-0225757.
 PR 14-AUG-2000: 2000US-0225758.
 PR 14-AUG-2000: 2000US-0225759.
 PR 18-AUG-2000: 2000US-0226781.
 PR 22-AUG-2000: 2000US-0226868.
 PR 22-AUG-2000: 2000US-0227182.
 PR 23-AUG-2000: 2000US-0227009.
 PR 30-AUG-2000: 2000US-0228924.
 PR 01-SEP-2000: 2000US-0229287.
 PR 01-SEP-2000: 2000US-0229343.
 PR 01-SEP-2000: 2000US-0229344.
 PR 05-SEP-2000: 2000US-0229509.
 PR 05-SEP-2000: 2000US-0229513.
 PR 06-SEP-2000: 2000US-0230437.
 PR 08-SEP-2000: 2000US-0231242.
 PR 08-SEP-2000: 2000US-0231243.
 PR 08-SEP-2000: 2000US-0231244.
 PR 08-SEP-2000: 2000US-0231413.
 PR 08-SEP-2000: 2000US-0231414.
 PR 08-SEP-2000: 2000US-0232080.
 PR 12-SEP-2000: 2000US-0231968.
 PR 14-SEP-2000: 2000US-0232397.
 PR 14-SEP-2000: 2000US-0232398.
 PR 14-SEP-2000: 2000US-0232399.
 PR 14-SEP-2000: 2000US-0232400.
 PR 14-SEP-2000: 2000US-0233063.
 PR 14-SEP-2000: 2000US-0233064.
 PR 14-SEP-2000: 2000US-0233065.
 PR 21-SEP-2000: 2000US-0234223.
 PR 21-SEP-2000: 2000US-0234274.
 PR 25-SEP-2000: 2000US-0234997.
 PR 25-SEP-2000: 2000US-0234998.
 PR 26-SEP-2000: 2000US-0235884.
 PR 27-SEP-2000: 2000US-0235884.
 PR 27-SEP-2000: 2000US-0235836.
 PR 29-SEP-2000: 2000US-0236327.
 PR 29-SEP-2000: 2000US-0236367.
 PR 29-SEP-2000: 2000US-0236368.
 PR 29-SEP-2000: 2000US-0236369.
 PR 29-SEP-2000: 2000US-0236370.
 PR 02-OCT-2000: 2000US-0236802.
 PR 02-OCT-2000: 2000US-0237037.
 PR 02-OCT-2000: 2000US-0237038.
 PR 02-OCT-2000: 2000US-0237039.
 PR 02-OCT-2000: 2000US-0237040.
 PR 13-OCT-2000: 2000US-0239935.
 PR 13-OCT-2000: 2000US-0239937.
 PR 20-OCT-2000: 2000US-0240960.
 PR 20-OCT-2000: 2000US-0241221.
 PR 20-OCT-2000: 2000US-0241785.
 PR 20-OCT-2000: 2000US-0241786.
 PR 20-OCT-2000: 2000US-0241787.
 PR 20-OCT-2000: 2000US-0241808.
 PR 20-OCT-2000: 2000US-0241809.
 PR 20-OCT-2000: 2000US-0241826.
 PR 01-NOV-2000: 2000US-0244617.
 PR 01-NOV-2000: 2000US-0244617.
 PR 08-NOV-2000: 2000US-0246474.
 PR 08-NOV-2000: 2000US-0246475.
 PR 08-NOV-2000: 2000US-0246476.
 PR 08-NOV-2000: 2000US-0246477.
 PR 08-NOV-2000: 2000US-0246478.
 PR 08-NOV-2000: 2000US-0246523.
 PR 08-NOV-2000: 2000US-0246524.
 PR 08-NOV-2000: 2000US-0246525.
 PR 08-NOV-2000: 2000US-0246526.
 PR 08-NOV-2000: 2000US-0246527.
 PR 08-NOV-2000: 2000US-0246528.
 PR 08-NOV-2000: 2000US-0246532.

PR 08-NOV-2000: 2000US-0246609.
 PR 08-NOV-2000: 2000US-0246610.
 PR 08-NOV-2000: 2000US-0246611.
 PR 08-NOV-2000: 2000US-0246613.
 PR 17-NOV-2000: 2000US-0249207.
 PR 17-NOV-2000: 2000US-0249208.
 PR 17-NOV-2000: 2000US-0249209.
 PR 17-NOV-2000: 2000US-0249210.
 PR 17-NOV-2000: 2000US-0249211.
 PR 17-NOV-2000: 2000US-0249212.
 PR 17-NOV-2000: 2000US-0249213.
 PR 17-NOV-2000: 2000US-0249214.
 PR 17-NOV-2000: 2000US-0249215.
 PR 17-NOV-2000: 2000US-0249216.
 PR 17-NOV-2000: 2000US-0249217.
 PR 17-NOV-2000: 2000US-0249218.
 PR 17-NOV-2000: 2000US-0249244.
 PR 17-NOV-2000: 2000US-0249245.
 PR 17-NOV-2000: 2000US-0249264.
 PR 17-NOV-2000: 2000US-0249265.
 PR 17-NOV-2000: 2000US-0249297.
 PR 17-NOV-2000: 2000US-0249299.
 PR 17-NOV-2000: 2000US-0249300.
 PR 01-DEC-2000: 2000US-0250160.
 PR 01-DEC-2000: 2000US-0250391.
 PR 05-DEC-2000: 2000US-0251030.
 PR 05-DEC-2000: 2000US-0251988.
 PR 05-DEC-2000: 2000US-0251989.
 PR 06-DEC-2000: 2000US-0251479.
 PR 08-DEC-2000: 2000US-0251856.
 PR 08-DEC-2000: 2000US-0251868.
 PR 08-DEC-2000: 2000US-0251869.
 PR 08-DEC-2000: 2000US-0251989.
 PR 11-DEC-2000: 2000US-0254097.
 PR 05-JAN-2001: 2000US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Barash SC, Ruben SM;
 WPI: 2001-465566/50.
 P-PSDB: AAU23090.
 Novel polypeptides and polynucleotides useful for diagnosis, preventing, treating neural, immune system, muscular, reproductive, pulmonary, cardiovascular, renal, proliferative disorders and cancerous diseases
 Claim 4: SEQ ID No 186: 1180pp; English.
 The present invention relates to the isolation of novel human enzyme polypeptides (AAU22915-AAU3814), and the cDNA and genomic sequences encoding them. The enzyme polypeptides of the invention may comprise the functional classes of oxidoreductases, transferases, hydrolases, lyases, isomerases or ligases. The sequences of the invention are useful in diagnosis, treatment, prevention and/or prognosis of a wide range of disorders including hyperproliferative disorders (e.g. cancer), immunodeficiency disorders (e.g. AIDS) autoimmune disorders (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis), blood-related disorders (e.g. haemophilia), reproductive disorders (e.g. infertility) and infectious disorders (e.g. influenza). The polynucleotides of the invention can also be used in gene therapy. AA40785-AA54164 represent cDNA sequences encoding for the novel human enzyme polypeptides of the invention.
 CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences.
 Sequence 1967 BP: 723 A: 344 C: 416 G: 476 T: 8 other:

Alignment Scores:

Prod. No.:	1.2e-252	Length:	1967
Score:	3176.00	Matches:	636
Percent Similarity:	97.11%	Conservative:	2
Best Local Similarity:	96.80%	Mismatches:	18
Query Match:	59.80%	Indels:	5
DB:	22	Gaps:	1

US-09-515-363C-2 (1-1025) x AAS40960 (1-1967)

```

QY 294 GLUAsnValAlaAlaArgLaseProGluLeuGlnLeuArgProGluMet 313
DB 3 GAAAGGTGGCAGCAGAGATCCCGGAGCCGAAAGTCCATCCATGAGGCTTACCAATG 62
QY 314 GLUValAlaGlnProAlaLeuGlnGlyLysAsnIleIleCysLeuProThrGlySer 333
DB 63 GAAGTGGCCGAGCCAGCTTGGAGGAAATATCATCTGCTCCCTACAGGAGGT 122
QY 334 GLYLeuThrArgValAlaValTyrIleAlaLysAspHisLeuAspLysLysLysAla 353
DB 123 GGAAGAACAGAGAGTGGCTGTTTACATTTGCCAGAGATCCTTAACAAAGAACAAAAAGCA 182
QY 354 SerGluProGlyLysValIleValLeuValAsnLysValLeuValGlnLeuPhe 373
DB 183 TGTGAGCTGGAGAAAGTTATAGTCTTGTGCAATAGGTAAGTCTTGTGAGTACGCTCTT 242
QY 374 ArgLysGluPheGlnProPheLeuLysLysTyrTyrArgValIleGlyLeuSerGlyAsp 393
DB 243 CGCAGAGAGTCCAGCCATTTTGAAGAAATGATGCTGTTTATGATTAAGTGGTGAAT 302
QY 394 ThrGluLeuLysIleSerPheProGluValValLysSerCysAspIleIleIleSerThr 413
DB 303 ACCCACTGAAATATCATCTTCCAGAAAGTTGTAAGTCTGTTATATTATATCATGACA 362
QY 414 AlaGlnIleLeuGlnAsnSerLeuLeuAsnLeuGlnGlnLysAspAlaGln 433
DB 363 GCTCAATTCCTTGAAGAACTCCTCTTAACCTTGAAGAAATGAGAGATGCTGCTGCTCA 422
QY 434 LeuSerAspPheSerLeuIleIleIleAspGlnCysHisHisThrAsnLysGlnVal 453
DB 423 TTGTGAGCTTTTCCCTCATATCATGATGATGATGATGATGATGATGATGATGATGATG 482
QY 454 TyrAsnAsnIleMetArgHisTyrLeuMetGlnLysLysLysAsnAsnIleGlnLys 473
DB 483 TATATATATCATGATGAGCTTATTTGATGACAGAGTGAAGAAACATATACCTCAAGAAA 542
QY 474 GluAsnLysProValIleProLeuProGlnIleLeuGlnLeuThrAlaSerProGlyVal 493
DB 543 GAAAGAACAGCAGTATCCCTCTCCATGATGATGATGATGATGATGATGATGATGATG 602
QY 494 GlyGlyAlaThrLysGlnAlaLysAlaGlnGlnHisIleLeuLysLysGlnAlaAsnLeu 513
DB 603 GGAAGGGCCGAGCAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 662
QY 514 AspAlaPheThrIleLysThrValLysGlnAsnLeuAspGlnLeuLysAsnGlnIleGln 533
DB 663 GATGCAATTTCTATTAAGTCTTAAAGAAACCTTGCATGATGATGATGATGATGATGATG 722
QY 534 GluProCysLysLysPheAlaIleAlaAspAlaThrArgLysAspProPheLysGlnLys 553
DB 723 GAGCAGTCAAGAAAGTTTGCATGATGATGATGATGATGATGATGATGATGATGATG 782
QY 554 LeuLeuGlnIleMetThrArgIleGlnIleThrCysGlnMetSerProMetSerAspPhe 573
DB 783 CTCTTCAAAATATGACAGAGATTAACAACTTATGTCAAATATGTCAAATATGTCAAAT 842
QY 574 GlyThrGlnProTyrGlnGlnIleProAlaIleGlnMetGlnLysLysAlaIleLysGly 593
DB 843 GGAAGTCAAGCCTATGACAGATGAGCCTATTAATGAAAAAAGCTGTAAGAGAAAGTA 902
QY 594 AsnArgLysGluArgValCysAlaGlnHisLeuArgLysTyrAsnGlnValLeuGlnIle 613
DB 903 AATCGCAAGAGAGTGTGTCGAGAAACATTTGAGAGATGATGAGGCTTACCAATTT 962

```

```

QY 614 AsnAspThrIleArgMetIleAspAlaTyrThrHisLeuGlnThrPheTyrAsnIleGln 633
DB 963 AATGACACAAATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1022
QY 634 LysAspLysLysPheAlaValIleGlnAspAspSerAspGlnGlyGlyAspAspThrTyr 653
DB 1023 AAGATTAAGCACTTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1082
QY 654 CysAspGlyAspGluAspGluAspAspLeuLysLysProLeuLysLeuAspGlnThrAsp 673
DB 1083 TGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1142
QY 674 ArgPheLeuMetThrLeuPhePheGlnAsnAsnLysMetLeuLysArgLeuValLysGln 693
DB 1143 AGATTTTCATGACTTATTTTAAAAAATTAATTAATTAATTAATTAATTAATTAATTAAT 1202
QY 694 ProGluTyrGluAsnGlnLysLeuThrLysLeuArgAsnThrIleMetGlnMetGlnThr 713
DB 1203 CCGAGATATGAAATGAAAGCTGACCAATTAATTAATTAATTAATTAATTAATTAATTAAT 1262
QY 714 ArgThrGluGluSerAlaArgGlyIleIlePheThrLysThrArgGlnSerAlaTyrAla 733
DB 1263 AGCAGTACAGAAATGACAGCAGAGAAATATCTTACAAAAACAGCAGAGATGATGATGATG 1322
QY 734 LeuSerGlnThrIleThrGlnAsnGlnLysPheAlaGlnValGlyValLysAlaHisHis 753
DB 1323 CTCTCCAGTGTGATTAAGTAAATGAAATTTGCTGAAAGTGAAGTGAAGTGAAGTGAAG 1382
QY 754 LeuIleGlyAlaGlyHisSerSerGluPheLysProMetThrGlnAsnGlnGlnGln 773
DB 1383 CTGATTTGAGAGTGGACAGCAGAGTGAATTAATTAATTAATTAATTAATTAATTAATTAAT 1442
QY 774 ValIleSerLysPheArgThrGlyLysIleAsnLeuLeuIleAlaThrThrValAlaGln 793
DB 1443 GTCATTAAGTAAATTTGCGACTGGAAGAAATTAATTAATTAATTAATTAATTAATTAAT 1501
QY 794 GlnGlyLeuAspIleLysGlnCysAsnIleValIleArgTyrGlyLeuValIleArgGln 813
DB 1502 GAAGGTGTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1561
QY 814 IleAlaMetValGlnAlaArgGlyArgAlaArgAlaAspGluSerThrTyrValGln 833
DB 1562 ATATCC-ATGCTGACGCC-GGTGGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1618
QY 834 AlaHisSerGlySerGlyValIleGlnHisGlnThrValAsnAspPheArgGlnGln 853
DB 1619 GCTCAGACAGTGTTCAGAGATTAATGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1678
QY 854 MetTyrLysAlaIleHisCysValGlnAsnMetLysProGlnGlnTyrAlaHisHis 873
DB 1679 ATGTATTAAGCTATTC-ATTGCTTCAATTAATTAATTAATTAATTAATTAATTAATTAAT 1735
QY 874 LeuGlnLeuGlnMetGlnSerIleMetGlnLysLysMetLysThrLysArgAsnIleAla 893
DB 1736 TTGAGATTAACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1795
QY 894 LysHisTyrLysAsnAsnProSerLeuIleThrPheLeuCysLysAsnGlnSerGlnGln 913
DB 1796 AAGCATTAACAGATTAACCATCACTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 1855
QY 914 AlaCysSerGlyCysLysPheHisValIleGlnLysMetHisHisValAsnMetThrPro 933
DB 1856 GCTGTGTTGGGAGATATTCATGAAATTAATTAATTAATTAATTAATTAATTAATTAAT 1915
QY 934 GluPheLysGlnLeuTyrIleValArgGlnAsnLysAlaLeuGlnLysLys 950
DB 1916 GAATTAAGAGACTTACATTTGGAAGAGAAAGAAAGCAAGCTGCCAAAGAGAA 1966

```

RESULT 5
AB04916
ID AB04916 standard; DNA; 1443 BP.
XX

AC ABA04916;
 XX
 DT 05-MAR-2002 (first entry)
 XX
 DE Human RNA helicase RH16 related DNA sequence *
 XX
 KW Human; RH16; RNA helicase; cytosolic; virucide; anti-HIV;
 KW immunosuppressive; immunostimulatory; antineoplastic; antidiabetic;
 KW antileukemic; osteoporosis; diabetes; but many others listed;
 KW antileukemic; cancer; infection; HIV; hepatitis; genetic disease;
 KW autoimmune disease; graft rejection; vaccine; ds
 XX
 OS Homo sapiens.
 XX
 PN WO200185955-A1.
 XX
 PD 15-NOV-2001.
 XX
 PF 11-MAY-2001: 2001WO-FR01441.
 XX
 PR 11-MAY-2000: 2000FR-0006030.
 XX
 PA (ISTA-) ISTAC.
 PA (INSP) INST PASTEUR LILLE.
 XX
 PI Bahr G, Cocude C, Capron A;
 XX
 DR WPI: 2002-082898/11.
 XX
 PT New polypeptide, useful for treating and diagnosing cancer or
 PT inflammation, and drug screening, comprises a human polynucleotide
 PT homologous to RNA helicase
 XX
 XX Example 2: Page 94-95; 114pp; French.
 XX
 CC The present invention relates to human RH16 (see AAM/798). RH16 is a
 CC 116kDa protein and has homology to RNA helicases (DEH1 box). RH16 and
 CC its coding sequence are useful for treating cancer acute or chronic
 CC infections (especially by HIV or hepatitis B or C); inherited genetic
 CC diseases; (auto)immune diseases (particularly rheumatism, arthritis,
 CC arteriosclerosis, osteoporosis and diabetes, but many others listed) and
 CC to prevent graft rejection. RH16 and its coding sequence are also useful
 CC for inducing, or increasing, the immune response to a vaccine. The
 CC present sequence was used in an example from the present invention.
 XX
 SO Sequence 1443 BP; 508 A; 261 C; 313 G; 360 T; 1 other:
 XX
 Alignment Scores:
 Pred. No.: 3 436-192 Length: 1443
 Score: 2442.00 Matches: 475
 Percent Similarity: 99.388 Conservative: 2
 Best Local Similarity: 98.968 Mismatches: 3
 Query Match: 45.988 Indels: 0
 DB: 24 Gaps: 0
 US-09-515-363c-2 (1-1025) x ABA04916 (1-1443)
 QY 174 GUAAGUANTTTPheseraAlaPheLeuAanValLeuArgGlnhrcjlyAsnAsnGlnleu 193
 DB 2 AAGAAAGAGTCTCTCTGATTTCTGAATGTTCTCTCTCAAAAGCAATGCAATTTATCA 61
 QY 194 ValGlnGlnLeuThrGlySerAspCysSerGlnSerAsnAlaGlnLeuGlnLeuSer 213
 DB 62 GTCCAAAGAGTAAAGAGCTCTGATTTCTCTCAAAAGCAATGCAATGCAATTTATCA 121
 QY 214 GlnValAspGlyProGlnValGlnGlnGlnLeuSerThrPhyValGlnProAsnGln 233
 DB 122 CAAGTTCATGCTCTCAATAGAGCAAGCAATCTTTCACACAGCTTCAACCAAAATTCG 181
 QY 234 GlnGlyGlnValTrrpGlyMetGlnAsnAsnSerSerGlnSerSerPheValAspSer 253
 DB 182 GAGAAAGAGAGTCTGGGCGCATGGAATTAATCATCATGATTCATCTTTCACATCTCT 241

QY 254 ValValSerGlnSerAspThrSerLeuAlaGlnGlySerValSerCysLeuAspThrSer 273
 DB 242 GTAGTTTCAGAAATCAGACACAGATTTCAGAGAGAGAGTCTGCTGCTTCAATTAAGT 301
 QY 274 LeuGlnHisAsnSerAsnMetGlySerAspSerGlyThrMetGlySerAspSerAspGlu 293
 DB 302 CTGGACATTAACAGCAACATGGGAGTATTCAGGACCATGGGAGGATTTCAATAGAA 361
 QY 294 GlnAsnValAlaAlaIleArgAlaSerProGlnProGlnLeuGlnLeuArgProGlnMet 313
 DB 362 GAGAAATGGCGAGCAAGAGCATCCCGGAGCGAGAACTCCAGCTCAGCTTACCAATG 421
 QY 314 GlnValAlaGlnProAlaLeuGlnGlnGlnValAsnGlnValLeuValGlnLeuPhe 333
 DB 422 GAAGTTCGCCACCCAGCTCTGGAGAGGAAATATCATCATCTGCTTCCATCAAGAGT 481
 QY 334 GlyThrThrArgValAlaValTrrpIleAlaIleAspHisIleAspLysValAspGln 353
 DB 482 GGAAGAAACAGAGTGGCTGTTTACCTTCCAGAGATCCTTAGACACAGCAAGAAAGCA 541
 QY 354 SerGlnProGlyLysValIleValLeuValAsnLysValLeuLeuValGlnLeuPhe 373
 DB 542 TCTGAGCTCGAAGAAATATAGTTCTGTCATATAGTACTAGTTGAACAGCTTTC 601
 QY 374 ArgLysGlnPheGlnProPheLeuLysLysTrrpArgValIleGlyLeuSerGlyAsp 393
 DB 602 CGCAAGAGAGTCCAAACCATTTTGAAGAAATGATCTGTTATTTGATTAAGTTCAT 661
 QY 394 ThrGlnLeuLysIleSerPheProGlnValValLysSerCysAspIleIleLeuThr 413
 DB 662 ACCCACTGAAATATTCATTTCCAGAACTTGCAAGTCTGATATTATTCATACAA 721
 QY 414 AlaGlnIleLeuGlnAsnSerLeuLeuAsnGlnLysGlnValGlnValGln 433
 DB 722 CCTCAATCTCTGAAAGTCTCTTAACTTGAAGTGAAGAGAGTCTGCTTCTCAAA 781
 QY 434 LeuSerAspPheSerLeuIleIleIleAspGlnGlnGlnGlnGlnGlnGlnGlnGln 453
 DB 782 TTGTAGACTTTTCTTATATTCATATGATTAATGATTCATCAACCAAGAAAGAAATG 841
 QY 454 TyrAsnAsnIleMetArgHisTrrpLeuMetGlnLysLeuLysAsnAsnArgLysLys 473
 DB 842 TATATATCATCATATAGGATTTATTTGATGCAAGATTTGAAACATATAGACTTAAGAA 901
 QY 474 GlnAsnLysProValIleProLeuProGlnIleLeuGlnLysLeuThrAlaSerProGlyVal 493
 DB 902 GAAACAAACACGATGATTCCTTCATGATGATGATGATGATGATGATGATGATGAT 961
 QY 494 GlnGlnAlaThrLysGlnAlaLysAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 513
 DB 962 GGAGGGGCCGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1021
 QY 514 AspAlaPheThrIleLysThrValLysGlnLysAsnGlnLeuLysAsnGlnIleGln 533
 DB 1022 GATGCATTTACTATTAAGATGTTTAAAGAAACCTTGATCAACTGAAACCAAAACACAG 1081
 QY 534 GlnProCysLysLysPheAlaIleAlaAspAlaThrArgGlnAspProPheLysGlnLys 553
 DB 1082 GAGCCATGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1141
 QY 554 LeuLeuGlnIleMetThrArgIleGlnThrLysGlnMetSerProMetSerAspPhe 573
 DB 1142 CTCTAGAAATATATGACAAAGATTAACCTTATTTGCAATGATGATGATGATGATGAT 1201
 QY 574 GlyThrGlnProGlnGlnGlnIleAlaIleGlnMetGlnLysLysAlaAlaLysLysGly 593
 DB 1202 GGAATCTCAACCTATGAAACATAGGGCATTTCAATAGGAAAAAGCGCAAAAGAAAGCA 1261
 QY 594 AsnArgLysGlnArgValLysAlaGlnIleLysLeuArgLysTrrpAsnGlnAlaLeuGlnIle 613
 DB 1262 AATCCAAAGAAAGAGTGTGTCAGACATTTGAGAGAGTACAAATAAAGCCCTTAATATT 1321
 QY 614 AsnAspThrIleArgMetIleAspAlaTrrpThrHisLeuGlnIleThrPheTrrpAsnGln 633

DB 919 TATACATTGTTCAAAATATGAAACAGAGAGTATGCTCAACGATTTCGAAATACCA 978
 OY 877 nmetglnserllemetglulyslysmetlthrylsalgaalealayslstryly 897
 DB 979 GATGCAAGATATATGCAAAAGAAATGAAACCAAGATCAAAATATGCAACCATTTACCA 1038
 OY 897 saanaaproserleullethrphleucylslyasnccssvalleuplacyssergl 917
 DB 1039 GAATTAACCATCTACTATACATCTTCTTGCAAAAACATGCAATGCTATCTGTTTCG 1098
 OY 917 ygluaspillehlsvalileglulysmethlshlsvalasnmcthrprophphelsgl 937
 DB 1099 GGAAGATATCCATGATGATATGAGAAATGCAATGCAATATACCCCAATATTCAGCA 1158
 OY 937 uleutryllevalarggluasnlysalaleuglnlyscysalaasplvrglnleas 957
 DB 1159 ACTTACATTTGTAGAGAAACCAAACTGCTGCAAAAGAGTGGCCGATATCAATAA 1218
 OY 957 ncllyglullellecgylscylglinalatrpqglythrmcmelvalhlslysglyle 977
 DB 1219 TGGTGAATCATCTGCAAAATGCGCCGCGCTGGGGAAATATGATGGTGCACAAAGCTT 1278
 OY 977 uaapleuprocylsleulysllearganphevalvalphuclysasnserthrlly 997
 DB 1279 AGATTGGCTTGTCTCAAAATAGCAATTTTGTAGTGGTTTAAATATATATTCACACA 1338
 OY 997 slsglnltyrlylstrpvalgluendprolethrhphleuasnleuasptrysergl 1017
 DB 1339 GAAACATATCAAAAGTGGTGAATTTACTTACATTCCTCAATCTTACTATTCACA 1398
 OY 1017 ucyscylsleupheseraspqluasp 1025
 DB 1399 ATGCTGTTATTTAGTATGATGAGAT 1423

RESULT 7
 AA161071
 ID AA161071 standard; cDNA; 1382 BP.
 AC AA161071;
 DT 22-OCT-2001 (first entry)
 DE Human polynucleotide SEQ ID NO 5060.
 XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 XX peripheral nervous system; neuropathy; central nervous system; CNS;
 XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chondralysis;
 XX chemokinesis; thrombolytic; drug screening; arthritis; inflammation;
 XX leukaemia; ss.
 OS Homo sapiens.
 XX
 PN MO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000MO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0596042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSEQ INC.
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao Q, Zhou P, Goodrich R, Dimanic RT;

XX WP1: 2001-442253/47.
 DR P-PSDB: AAM41915.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Claim 1: SEQ ID NO 5060; 10078bp; English.
 CC
 CC The invention relates to human nucleic acids (AA157798-AA161469) and
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemia and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 SQ Sequence 1382 BP: 513 A; 217 C; 285 G; 367 T; 0 other:
 XX
 Alignment Scores:
 Pred. No.: 1,29e-165 Length: 1382
 Score: 2120.00 Matches: 407
 Percent Similarity: 98.32% Conservative: 2
 Best Local Similarity: 97.84% Mismatches: 5
 Query Match: 39,92% Indels: 2
 DB: Gaps: 0

US-09-515-363C-2 (1-1025) x AA161071 (1-1382)
 OY 612 Glnlleasnaspthrleargmetlleaspaalatyrthrhlslsleugluthrphleuasn 631
 DB 13 CAAATTAATGACACAAATTCGATGATGATGATGATGATGATGATGATGATGATGATGAT 72
 OY 632 Gluglulysaspplyslysphealavallleulaspaspserasppluglylylaspasp 651
 DB 73 GAAAGGAAAGATMAAGAAATTTGCACTCATAGATGATGATGATGATGATGATGATGAT 132
 OY 652 Glutrycysaspplaspaspplaspaspplaspaspplaspaspplaspaspplaspasp 671
 DB 133 CAGTATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 142
 OY 672 Thrasp-argpheleumetrhleuphephegluasnaslysmet-leulsaaleua 691
 DB 193 ACAGATACGATTTCTCATGACTTTATTTTGAACCAATTAATAGTTGAAATAGTCAAT 252
 OY 691 lagluasnproglutylrglunsglulysleuthrlyleuargasnthrllwctug 711
 DB 253 CTGAACACCCCTTAATATGAAATGAAAGCTGACCAATTAAGAAATCCCAATATAGC 312
 OY 711 lntythrargthrnglulseralarglyllelephetrhllystrhrnglulsera 731
 DB 313 AATATACTAGAGTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 372
 OY 731 latyralaleuserglntrpillethrcluasnglulysphealaglvalcylvalysa 751
 DB 373 CATATGCGCTTCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 432
 OY 751 lahlshlsleullelglalaglhlhsserserglupheylspromethrnglunsgl 771
 DB 433 CCCACCATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 492
 OY 771 lnllyglulvallelserlyspheargthrnglyllylleasluleullelathrlthry 791
 DB 493 AAAAAGAGTCAATTGTAATTTGCGACTGCAAAATTAATATGCTTATCGTTACACAG 552

QY	791	alaiaaglugluglyleuaspiilelysylglucysasnilevallelkrqtyrlyleuvalt	811
Db	553	TCGCAGAGAAAGGCTCGATATTAAAGAAATGTAACTTCCTTATCCGCTTAIACCTCGTCA	612
QY	811	hrasngluilealmetvalglnalaaraglyargalaarvalaaspheerthryr	831
Db	613	CCAATGAATAATAGCCATGTGTCACAGGCCCGTGTGCAGCCAGACTGATGAGACACCTACG	672
QY	831	alleuvalalahlsserglyserglyvalillegluhiscuhyrvalasnsprearag	851
Db	673	TCCCGGTGGTCACAGTGGTTCAGGAGTTATGAAACATGACACATTAATCATTCGACG	732
QY	851	lutysmetctyrrlysalailehiscysvalglnasmetcysprogluhyrvalah	871
Db	733	AGAAAGATGATGTATAAAGCATATCATTTGCTTCMAATTAAGAACAGACAGATGTGTC	792
QY	871	lslysileleuglualeuglnmetglnserilemetgludyslysmetlysthrlysarqa	891
Db	793	ATAAGATTTTGGAAATTACACATGCAAGATATATATGAAAGAAATGAATACCAAGACAGA	852
QY	891	snllealalyshistyrlyrsasnasnpproserleuilethrheleucyslysasrcyss	911
Db	853	ATATTGGCAAGCATTTACCAAAATAAACCCATCATATAATCTTCTTTGMAAAACTGCA	912
QY	911	ervalleuvalacysserglygluasplilehisevalilegludysmethisusvalasnm	931
Db	913	GTCGGCTACGCTGTTCGGGGAGAGATATCATGTATTAATGAAATGCATACGTCAATA	972
QY	931	etthrprogluphelrysglualeuylrilevalargsluashlysalaenluhytlysc	951
Db	973	TGACCCCGAATTCAGGAACCTTTACATGTTAAGAGAAATCAACACACTGTMAAAGACCT	1032
QY	951	ysalaaspyrrglnileasnlgyluilelleecyslyscyslylglalstirpglytrtm	971
Db	1033	GTCGGCGACTTCACAAATMAAGTGGAAATCATCTGCAAAATGGCCAGGCTGGGGAAACA	1092
QY	971	etmetvalhislysglyleuaspleuprocysleuyslileargsnphevalvalalp	991
Db	1093	TGATGGTCGACMAAGGCTTGATTTGCTTCCTTCGTCAAAATTAAGCAATTTCTAGTGCTT	1152
QY	991	helrysasnasnserthrlyrslysglntryllysllystirpvalsglualeupnilethrph	1011
Db	1153	TCAAAATAATATTCACAAAGAAACAAATACAAAAAGTGGTGAATTAACATACATATTC	1212
QY	1011	roasleuasprryserrglucycysleupheserrspoluasp	1025
Db	1213	CCAAATCTTGACTATTCAGAAATGCTGTATTATTAATGATAGAGAT	1256
RESULT 8			
ABAO4913	ID	ABAO4913 standard; DNA; 1284 bp.	
XX	XX	ABAO4913:	
AC	XX	05-MAR-2002 (first entry)	
DT	XX	Human RNA helicase RH16 related DNA sequence #2.	
DE	XX		
XX	XX	Human: RH16; RNA helicase; cytosolic; virucide; anti-HIV;	
KM	XX	Immunosuppressive; Immunostimulatory; antitubercular; antilarthritic;	
KM	XX	antiartherosclerotic; osteoprotective; antidiabetic; hepatotropic;	
KM	XX	antiinflammatory; cancer; infection; HIV; hepatitis; genetic disease;	
KM	XX	autoimmune disease; graft rejection; vaccine; ds.	
OS	XX	Homo sapiens.	
XX	XX	MO200185955-A1.	
PN	XX	15-NOV-2001.	
PD	XX		
XX	XX	11-MAY-2001.2001MO-FR01441.	
FE	XX		
XX	XX		

PR	11-MAY-2000; 2000FR-0006030.
XX	
XX	(ISTA-) ISTAC.
PA	(INSP) INST PASTEUR LILLE.
PA	
XX	
PI	Bahr G, Cocude C, Capron A;
XX	
DR	WPI; 2002-082898/11.
XX	
PT	New polypeptide, useful for treating and diagnosing cancer or
PT	inflammation, and drug screening, comprises a human polynucleotide
PT	homologous to RNA helicase
XX	
PS	
XX	Example 2; Page 93-94; 114pb; French.
XX	
CC	The present invention relates to human RH16 (see AAM47798). RH16 is a
CC	116kDa protein and has homology to RNA helicases (DEX box). RH16 and
CC	its coding sequence are useful for treating cancer: acute or chronic
CC	infections (especially by HIV or hepatitis B or C); inherited genetic
CC	diseases; (auto) immune diseases (particularly Rheumatism, arthritis,
CC	arteriosclerosis, osteoporosis and diabetes, but many others listed) and
CC	to prevent graft rejection. RH16 and its coding sequence are also useful
CC	for inducing, or increasing, the immune response to a vaccine. The
CC	present sequence was used in an example from the present invention.
XX	
XX	Sequence 1284 BP; 489 A; 198 C; 261 G; 335 T; 1 other;

Protein No.:	2_48e-155	Length:	1284
Score:	1995.00 <td>Matches:</td> <td>378</td>	Matches:	378
Percent Similarity:	99.47%	Conservative:	0
Best Local Similarity:	99.47%	Mismatches:	2
Query Match:	37.56%	Indels:	0
DB:	24	caps:	0
US-09-515-363C-2 (1-1025) x ABA04913 (1-1284)			
QY	646	AspGIUGLIGYLAASPAPGILUTTCYASAPGILYSPGILUASPGLUASPAPLEULYS	665
DB	2	GATAGGGGTGGTGGATGATGAGTATTGGATGGGTATGAACGATGGATGATTTAACAAA	61
QY	666	ProLeuLYSLUASAPGILUTHASPARPHELEUMETHIRLEUPHEPEJLUASASULYS	685
DB	62	CGTTGAAACGCGATGAAACGATAGATTCTCATGACTTATTTT3AAACAAATAA	121
QY	686	MetLeuLYSATGLUeALAGLUASPRGILUTYGLUASNGILUYSLEUTHIRLYSLEATY	705
DB	122	ATGTTGAAAAGCGCGCTGAAAACCCAGAAATATGAAAATGAAAAGCGACCAATTAAGA	181
QY	706	AsnTHRILLEMETGUGINTYTRTHAFTGTHGUGLUSERALAARGYLIELLEPHETHR	725
DB	182	AATACCAATATGAGCAATATAGTATAGAGTACAGGAATCAGACAGGAGATATCTTACA	241
QY	726	LysTHRATGGLINSEALATATYALALEUSERGLINTPRILETHRGUANGILUYSRPHALA	745
DB	242	AAAACACACAGAGTGCATATAGCCCTTCCAGTGGATATCTGAAATGAAAAATTGGT	301
QY	746	GLUVALIGLYVALILYSALAHSHISLEULLEGLYLAAGLYNISSESERGLUPHELYSTRO	765
DB	302	GAAAGAGAGAGTCAAAAGCCCAACCATCTGATTGGAGCTGGACACAGCAGTAGTTCAAAACC	361
QY	766	MetTHRGLINANGILUGILINLYSGILUVALILESERLYSPHEATGTHRGILYSLILEASLEU	785
DB	362	ATGACACAGCAATGAAACAAAAGAAAGTATTAGTAATTTGGCACTGGAAATAATAACTG	421
QY	786	LeuLEUALATHRTHVALALAGLIGLUGLYLEUASPIILEYSGILUCYASANILEVALIIC	805
DB	422	CTTATCGGTACCACAGTGGCCGAGGAAGAGCTGTGATTTAAAGAAATTATCATTTGTAIC	481
QY	806	ARGTYRGILYLEUVALITHRASNGILUeALAMELYALGNALAAARGLYATGALAAATGALA	825
DB	482	CGTTATGGTCTCGTCCACCAATGAAGTAAGCCATGGTCCAGGCGCCGTGGTGAAGCCAGAGCT	541

QY 826 AspGluSerThrValValLeuValAlaHisSerGlySerGlyValIleGluHisGluThr 845
 DB 542 GATGAGGACCTACCTACCTGCTGCTCAGAGGTTCACAGCTTACGACCTTACGACCTGACGAC 601
 QY 846 ValAsnAspPheArgGluLeuMetMetTyrLeuAlaIleHisTyrValGluAsnMetLys 865
 DB 602 GTTAATGATTCGAGAGGAAGATGATTAAGCTATACATTTCTGTCMAATATACAA 661
 QY 866 ProGluGluTyrAlaHisLysIleLeuGluLeuGluMetGluHisSerIleMetGluLys 885
 DB 662 CAGAGAGATGCTCATTAAGATTTGGAATTGACATGCAAGTAAAGTAAAGCAAA 721
 QY 886 MetLysThrLysArgAsnIleAlaLysHisTyrLysAsnAspProSerLeuIleThrPhe 905
 DB 722 ATGAAAACCAAGAAATATTCGCCAGCATTTACAGAAATACCATCCTCAATACCTTTC 781
 QY 906 LeuCysLysAsnCysSerValLeuAlaCysSerGlyGluAspIleHisValIleGluLys 925
 DB 782 CTTTGCAAAAACGTCAGTGTGCTAGCTGTCTGGGCAAGATATTCATGATTAATGAGAA 841
 QY 926 MetHisHisValAsnMetThrProGluPheLysGluLeuTyrIleValArgGluAsnLys 945
 DB 842 ATGCATCAGCTCATATGACCCGCAATCCAGAACTTAAATGTAAACAAACAA 901
 QY 946 AlaLeuGluLysLysCysAlaAspTyrGluIleAsnGlyGluIleLeuCysLysCysGly 965
 DB 902 ACACGTCAAAAGAAAGTGTGCGCTATCAATTAATGAGTAAATCAATCTGAAATGTGCG 961
 QY 966 GluAlaIleTyrPheMetMetValHisLysGlyLeuAspIleProCysLysLysLeuArg 985
 DB 962 CAGCGTGGGCAACATGATGATGTCACAAAGGCTTGAATTTGCTGTCTCAAAATTAAG 1021
 QY 986 AsnPheValValPheLysAsnAsnSerThrLysLysGluTyrLysLysTyrPheValGlu 1005
 DB 1022 AATTTGTAGTGTGTTCAAAATTAATTCACAAAGAAACAAATACAAAAAGTGGTAGAA 1081
 QY 1006 LeuProIleThrPheProAsnLeuAspTyrSerGluCysCysLysPheSerAspGluAsp 1025
 DB 1082 TTACGATATCATCTCCCAATCTTGACTATTCAGAAATGCTGTTATTTATGATGAGAT 1141
 RESULT 9
 AAS41517
 ID AAS41517 standard; cDNA; 1319 BP.
 AC AAS41517;
 XX 17-DEC-2001 (first entry)
 DT
 DE cDNA encoding novel human enzyme polypeptide #713.
 XX
 KW Human: oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
 KW ligase; hyperproliferative disorder; immunodeficiency disorder;
 KW autoimmune disorder; neurological disorder; metabolic disorder;
 KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
 KW blood-related disorder; infectious disorder; gene therapy; cytostatic;
 KW anti arthritic; nephrotropic; anticoagulant; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO20015301-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01239.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225457.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.

PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0244674.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249219.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249266.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256179.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR MPI: 2001-465566/50.
 DR P-PSDB: AAU23647.
 XX
 PT Novel polypeptides and polynucleotides useful for diagnosing,
 PT preventing, treating neural, immune system, muscular, reproductive,
 PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
 PT diseases -
 XX
 PS Claim 4: SEQ ID NO 743; 1180bp; English.
 CC The present invention relates to the isolation of novel human enzyme
 CC polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences
 CC encoding them. The enzyme polypeptides of the invention may comprise the

CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
 CC isomerases or ligases. The sequences of the invention are useful in the
 CC diagnosis, treatment, prevention and/or prognosis of a wide range of
 CC disorders, including hyperproliferative disorders (e.g. cancer),
 CC immunodeficiency disorders (e.g. AIDS), autoimmune disorders
 CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
 CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
 CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
 CC blood-related disorders (e.g. haemophilia), reproductive disorders
 CC (e.g. infertility) and infectious disorders (e.g. influenza). The
 CC polynucleotides of the invention can also be used in gene therapy.
 CC AAU0785-AAU41684 represent cDNA sequences encoding for the novel human
 CC enzyme polypeptides of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.
 CC
 SQ Sequence 1319 BP; 491 A; 217 C; 294 G; 314 T; 3 other:
 XX
 Alignment Scores:
 Pred. No.: 156-152 Length: 1319
 Score: 1961.50 Matches: 406
 Percent Similarity: 93.64% Conservative: 6
 Best Local Similarity: 92.27% Mismatches: 17
 Query Match: 36.93% Indels: 16
 DB: 22 Gaps: 4
 US-09-515-363c-2 (1-1025) x AAU41517 (1-1319)
 QY 442 ILEASPLUCYSHSHISFTHASNLGSLAVALTYFASNSN11METAFHISLYR 461
 DB 23 GTTGATGATGATCATCACCAACAAAGAGAGTGATATATACATCATGAGGACAT 82
 QY 462 LEUMETGLNLSLEULYSASNSANRGLLEULYSLSGLIASNLSYPROVAL11PROLEU 481
 DB 83 TTGATGCAGACAGCTTGAAAAACATAGCTCAAGAAAGAAACAAACAGTGTCCCTT 142
 QY 482 PROGLN11LELGLYLEULHRAISERPROGLYVALIGLYAL1ATHRLYSGLN11ALS 501
 DB 143 CCTGAGATGCTGGAGTACAGCTTCCCTGTTGGAGGGGCCACGAGAGCAACCTAA 202
 QY 502 ALAIGLGLHSH11LEULYSLEUCYSALASNLLEULASPLA11PHETHR11LEULYS 521
 DB 203 GGTGAAGAACACATTTTAAACTATGTCCATCTTGATGATTTACATTAACCTGT 262
 QY 522 LYSGLIASNLSLEULYSASNSN11LEGLNGLUPROCYLSLSYSPHEAL11LE 541
 DB 263 AAAAGAAACCTGATCACTGAAAAACCAATACAGAGCCATGCAGAGATTGCCAT 322
 QY 542 ALAASPLA11THARGGLIUSP11PROPHLEYSGLIUSLEULGL11EMETHRA11LE 561
 DB 323 GCGATGACACACGAGACAGATCCATTTTAAAGGAAACCTTGAAATTAATGCAAGCAT 382
 QY 562 GLNTHRTYRCYSLN11METSERFOMETSERASPHEGLYTHRLN11PROTYRGLN11TR 581
 DB 383 CAAACTTATTTGCAATGAGTCCAAATGTCAGATTTGGAAATCAACCCATGACAA11GG 442
 QY 582 ALA11GLN11METGLIUSLSYLSAL11ALYSLSGLYAS11NARGLYGLIARGLYALS 601
 DB 443 GCCATTCAAATGAGAAAAAAGCTGCAAAAAGAGAAATCGCAAGAGAGTCTT11T32A 502
 QY 602 GLNHL11LEULGLYSYTRFNSGLI11AL11GLN11LEAS11ASPT11LEARGMET11LSP 621
 DB 503 GAAACATTTGAGAGAGTACAGAGGCTTCAAAATTAATGACACACATTTGATGATAC 562
 QY 622 ALATYTRH11SHLEUL11THRPHETRYASNGIUSGLYASPLYSYPHEA11AVALL1E 641
 DB 563 GCGTATACATCTTGAACCTTTCTATATGAAGAAAGATTAAGAGT11TGCAG11CAT 622
 QY 642 GLIUSPASPSERASPGLIUSGLYLSYASPSGLIUTRYCYASPSGLYASPSGLIUSPS 661
 DB 623 GAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 682

OY 662 AspleuLysProLeuLysLeuAspGluThrAspArgHisLeuMetThrLeuPhePro 681
 DB 683 GATTAAACAACTTGAACCTGGATGAAACAGATGATTTCTCATGATTTATTTT 742
 OY 682 GluAsnAspLysMetLeuLysArgLeuAlaGluAsnProGlyTyrGluAsnGluLysLeu 701
 DB 743 GAAACAAATTAATTTGTTGAAAGCGCTGGTGAACCCCAATATATGAAATGAAAGCTG 802
 OY 702 ThrLysLeuArgAsnThrIleMetGluGlnTyrThrArgThrGluGluLysAlaArgGly 721
 DB 803 ACCAAATTAAGAAATACCAATATGAGCAATATAGTACAGTACAGCAATATACACAGCA 862
 OY 722 IleIlePheThrLysThrArgGlnSerAlaTyrAlaLeuSerGlnThrIleThrGluAsn 741
 DB 863 ATTAATCTTACAAAACACAGACAGATGATGCTTCCCTGACAGGATTAAGTAAAT 922
 OY 742 GluLysPheAlaGluValGlyValLysAlaHisLeuIleGlyValAlaHisSerSer 761
 DB 923 GAAAAATTTGCTGACAGTACAGAGTCAAGCCCACTCTATATGACACTGACACAGACT 982
 OY 762 GluPheLysProMetThrGlnAsnGluGlnLysGluValLysSerLysArgThrGly 781
 DB 983 GATTCAAAACCATGACACAGATGAAACAAAAGAGTCAATAGTAAATTCGCACTGCA 1042
 OY 782 LysIleAsnLeuLeuIleAlaThrThrValAlaGluGluLysLeuAspIleLysGly 801
 DB 1043 AAAAAATCTGCTTATCTCTCC-ACAGTGGCAGAAAGAAATCTGTGATATAAGCATCT 1101
 OY 802 AsnIleValIleArgTyrGlyLeuValThrAsnGluIleValMetValGlnAlaArgGly 821
 DB 1102 AACATTGKATTCGTTATGCTCTGCTCCACCAATGAAATGCC-ATGCTCCAGCC-CGTGCT 1159
 OY 822 ArgAlaArgAlaAspGluSerThrTyrValLeuValAlaHisSerGlyValAlaIle 841
 DB 1160 CGAGTAGAGCGTATGAGAGCAACCTACGCTGCTGCTGCTA--CAGTGGTACGAGTTAC 1216
 OY 842 GluHisGluThrValAsnAspPheArgGluLysMetMetLysAlaIleHisCysVal 861
 DB 1217 GACTTGAACAGCTA---TGATTTCCGAGAGAA-----TGATGTATA 1254
 OY 862 GluAsnMetLys-----ProGluGluTyrAlaHisLysIleLeuGlnLeuGlnMet 878
 DB 1255 AGTATCATGCTCAATATATACCGAGAGGTT-GCTTATACCAATTTGGATACGATG 1313
 RESULT 10
 AAH15201
 ID AAH15201 standard; cDNA: 2613 BP.
 AC AAH15201:
 DT 26-JUN-2001 (first entry)
 XX Human cDNA sequence SEQ ID NO:13298.
 XX Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
 OS Homo sapiens.
 PN EP1074617-A2.
 PD 07-FEB-2001.
 XX 28-JUL-2000; 2000EP-0116126.
 PF 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX (HFLI-) HELIX RES INST.
 PA Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI; 2001-318749/34.
 DR
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 PS
 XX
 XX Claim 8: SEQ ID 13298; 2537bp + CD ROM; English.
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 3'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises at least 15 nucleotides, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
 CC AAH93893 represent human amino acid sequences; and AAH1629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SQ Sequence 2613 BP; 591 A; 737 C; 783 G; 502 T; 0 other;
 Alignment Scores:
 Pred. No.: 3,86e-101 Length: 2613
 Score: 1344.00 Matches: 291
 Percent Similarity: 59.55% Conservative: 136
 Best Local Similarity: 40.59% Mismatches: 238
 Query Match: 25.31% Indels: 52
 DB: Gaps: 12
 US-09-515-363C-2 (1-1025) x AAH15201 (1-2613)
 OY 306 LeuGlnLeuArgProTyrGluMetGluValAlaGlnProAlaLeuGluLysAsnIle 325
 DB 224 ATGGAGCTTGCTGCTCTCAACAAAGGAGGTGATCATGCTGCTGAGGAGCAAAATATC 283
 OY 326 IleIleCysLeuProThrGlySerGlyLysThrArgValAlaValTyrTleAlaLysAsp 345
 DB 284 ATCATCTGGCTGGCCAGCGGTGCCGGAAGACCGCGGCTGCTTATAGTGGCAACGGG 343
 OY 346 HisLeuAspLysLysLysLysAlaSerGluProGlyLysValIleValLeuValHisLys 365
 DB 344 CACCTAGAG-----ACTGTGATGAGAGCCAGGTGCTGATTTGCTACACAGG 391
 OY 366 ValLeuLeuValGluGlnLeuPheArgLysGluPheGlnProPheLeuLysTyrTyr 385
 DB 392 GTGACCTGGTGACCCAG---CATGTGTAAAGAGTTCCAGGCCCATGCTGATGACGATGG 448
 OY 386 ArgValIleGlyLeuSerGlyAspThrGlnLeuLysIleSerPheProGluValValLys 405
 DB 449 ACCGTGACACCCCTGAGTGGGAGATGGACACACTGCTGCTGGCCACATGATCCGCG 508
 OY 406 SerCysAspIleIleIleSerThrAlaGlnIleLeuGlnAsnSerLeuLeuAsnLeuGlu 425
 DB 509 TGCCATGACCTGCTCATCTGCACACAGAGCTTTCGAGATGAGGACATGACACGCGGAG 568
 OY 426 AsnGlyLysAspAlaGlyValGlnLeuSerAspPheSerLeuIleIleLysP2Lys 445
 DB 569 -----GAGGAGAGCAGCCTGAGCTGCTCTTCTCTGATGCTGATGATGATGCTG 622

[illegible]

OY	805	ILITRTGTYGLYLEUVALTHIRASNGULILEALMEVVALGIALATRGUYTVALEIAHQ	824
Db	1563	GTCGGCTTATGGGCTCTTGACCAATTGAATTTCCATGTGTCAGGCCAAGGGCGCTATCTJG	1642
OY	825	AIAAspGuseTrThrTyValLeuValAlAHISserGlySerGlyValIleGThrHisHu	844
Db	1643	GCCGATCAGAGTGTATACGGGTTGTATAGCAAACTGAAGGTAAACC GGAGACTGAAAGTGGAG	1702
OY	845	ThrValAsnAspPheArgGluLysMetMetTyrLysAlaIleHisCysValIleAsnMet	864
Db	1703	CTGATCAACGAGGGCTGTGGAGAGCGTAATGTAGCACAGCATGGCTGCTGTGCAGAATAAG	1762
OY	865	LysProGluGluTyrAlaHisLysLileLengLueLincMecGlnSerLileMetLiu	884
Db	1763	GACCAAGCCCGAGTACAGAGCCAAAGATTCGGGATCTCCACAGGCAACCTTGACCAAGCGG	1822
OY	885	LysMetLysThrLysArgAsnIleAlaLysHisTyrLysAsnAspProSerLeuIleThr	904
Db	1823	GGCGGCCAGAGCCAGCCGGGAGAACCCAGGCGACACTTCCAGTAGGACCACTGGAG	1882
OY	905	PheLueCysLysAsnCysSerValLeuAlaCysSerGlyGluAspIleHisValIleGlu	924
Db	1883	CTACTCTGCATCACTCATGCTATGGTGTGGCTGTGGCCATGGCAGACCACTGCCGAAGTGGAG	1942
OY	925	LysMetHisHisValAsnMetThrProGluPheLysGluLeuTyr---LysValArgGlu	943
Db	1943	GGCAACCCACCATGTCAATGTGAACCCCACACTTTCGAACTACTATATATGCTCCGAAAGAT	2002
OY	944	AsnLysAlaLeuGlnLysLysCysAlaAspTyrGlnIleAsnGlyGluIleLysLys	963
Db	2003	CCTGGTGCATCAACAAGATCTTCACAGGACTGGAAACCTGGGGGTCTCATAGCTATGAGT	2062
OY	964	--CysGlyGlnAlaTatPglTyrThrMetMetValHisLysGlyLeuAspLeuProCysLeu	982
Db	2063	AACTGTGGGGAGTCTGGGCTGTGCATGATATCTACAAGACGATGACGTCCGCAAGTCTC	2122
OY	983	LysLleArgAsnPheValValValPheLysAsnAsnSerThrLysLysGlnTyrLysLys	1002
Db	2123	AAAGTCCGCAC-----ATGCTCTGTGAACACCCCTCAGGGGGGATGCCAGCCAAAAG	2176
OY	1003	TryValGluLeuProIleThrPheProAsnLeuAspTyrSerGluCysCys	1019
Db	2177	TGGTCCGGCTGCCCTTCTCCGTGCCCTGACTTTGACTTCTCTGAGCATTTGT	2227
<hr/>			
RESULT 11			
ID	AA501149		
XX	AA501149 standard; DNA; 1258 BP.		
AC	AA501149:		
DT	12-SEP-2001 (first entry)		
DE	Interferon induced nucleic acid, IFN4.		
XX			
KW	Interferon induced nucleic acid, autoimmune disease; lupus erythematosus;		
KW	immunodeficiency; stroke; acquired immunodeficiency syndrome; AIDS;		
KW	graft rejection; viral infection; hepatitis; aplastic anaemia; cancer;		
KW	human immunodeficiency virus; HIV; immune-mediated glomerulonephritis;		
KW	haematologic disease; chronic neutropenia; myocardial infarction;		
KW	neurological disease; Alzheimer's disease; Parkinson's disease; tumour;		
KW	amyotrophic lateral sclerosis; spinal muscular atrophy; hemorr; IFN4; ds.		
OS	Homo sapiens.		
XX			
PH	Key	Location/Qualifiers	
FT	CDS	1..915	
FT	/tag= a		
FT	/product= "IFN4"		
XX	/note= "Interferon induced polypeptide"		
PN	MO200118208-A2.		
PD	15-MAR-2001.		

XX 08-SEP-2000; 2000MO-US24704.
 PF 08-SEP-1999; 99US-0152921.
 PR 20-OCT-1999; 99US-0160575.
 PR 20-JAN-2000; 2000US-0177104.
 PR 07-SEP-2000; 2000US-0656633.
 XX (CURA-) CURAGEN CORP.
 PA (BIO) BIOGEN INC.
 XX
 PI Peyman JA, Da Silva A, Hochman P, Hsu A;
 DR WPI: 2001-235201/24.
 DR P-PSDB: AA000286.
 XX
 PT New interferon induced polypeptides and polynucleotides, useful for the
 PT diagnosis, prevention and treatment of immunological, cell
 PT proliferative disorders, such as lupus erythematosus, cancer, stroke
 PT and Alzheimer's disease
 XX
 PS Claim 9; Page 29-32; 134pp; English.
 XX
 CC The sequence represents interferon induced nucleic acid IFN6. IFN
 CC nucleic acids and polypeptides are useful for treating or preventing a
 CC pathology associated with IFN polypeptide in a human. They are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of IFN polypeptide or polynucleotide. IFN nucleic
 CC acids, polypeptides and antibodies are useful for diagnosis, prevention
 CC or treatment of variety of immunological and cell proliferative
 CC disorders, such as autoimmune diseases e.g. lupus erythematosus,
 CC immunodeficiency diseases such as acquired immunodeficiency syndrome
 CC (AIDS), graft rejection, viral infections including hepatitis and human
 CC immunodeficiency virus (HIV), immune-mediated glomerulonephritis,
 CC haematologic diseases such as aplastic anaemia and chronic neutropenia
 CC and cancer. In addition they are also useful for treating or diagnosing
 CC various disorders associated with cell death, including myocardial
 CC infarction, stroke, neurological diseases including Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis and spinal muscular
 CC atrophy. IFN nucleic acids and polypeptides are also useful for
 CC identifying interferon-like proteins and interferon agonists, for
 CC screening drugs and compounds which inhibit or enhance IFN activity or
 CC function and as targets for the identification of small molecules that
 CC are immunostimulatory, immunosuppressive, or stimulate or suppress normal
 CC cell or tumour cell growth in mammals, including humans.
 XX
 SQ Sequence 1258 BP; 399 A; 267 C; 276 G; 316 T; 9 other:
 Alignment Scores:
 Pred. No.: 8,556-90 Length: 1258
 Score: 1201.50 Matches: 242
 Percent Similarity: 94.53% Conservative: 0
 Best Local Similarity: 94.53% Mismatches: 1
 Query Match: 22.62% Indels: 14
 DB: 22 Gaps: 1
 US-09-515-363c-2 (1-1025) x AAS01149 (1-1258)
 QY 1 MetSerAsnGlyTyrSerThrAspGluAsnPheArgTyrLeuIleSerCysPheArgAla 20
 DB 179 AATCGCAATGGGATATCCACAGACGAGAAATTCGGCTATCTATCTCGTCTTCAGAGGCC 238
 QY 21 ArgValLysMetTyrTyrIleGlnValGluProValLeuAspTyrLeuThrPheLeuProAla 40
 DB 239 ACGGTGAAAATGTACTTCA----- 258
 QY 41 GluValLysGluGlnIleGlnArgThrValAlaThrSerGlyAsnMetGlnAlaValGln 60
 DB 259 GAGGTGAAGGACGACATTTAGAGAGACAGTCCGACCTCGGAGACATGATGCCAGTTGAA 318
 QY 61 LeuLeuLeuSerThrLeuGluLysGlyValTyrPheLysGlyTyrPheArgGluPheVal 80
 DB 319 CTGCTGCTGACGACCTTGAGAAAGGAGTCTGCACTTGCTGGACTGCGAAATTCTG 378

QY 81 GluAlaLeuArgArgThrGlySerProLeuAlaIleArgTyrMetAsnProGlnLeuThr 100
 DB 379 GAGGCCCTCCGGAGAACCGGACGCTCTGGCCGCCCTACATGAACCTTCAAGCTCAG 438
 QY 101 AspLeuProSerProSerPheGluAsnAlaHisAspGlyTyrLeuGlnLeuLysLeu 120
 DB 439 GACTTGCCCTCTCCATCGTTTGAGAACCGCTATGATGATATCTCCAACTGTTCAACTG 498
 QY 121 LeuGlnProThrLeuValAspLysLeuValArgAspValLeuAspLysCysMetGlu 140
 DB 499 CTTCAGCCACACTCTGTGTGACAAAGCTTACTGTTACAGACGTTTGTGATAGTATGAG 558
 QY 141 GluGluLeuLeuThrIleGluAspArgAsnArgIleAlaAlaIleGluAsnAsnIlyAsn 160
 DB 559 GAGGAACCTCTGACATTCAGACAGCAAAACCGCATTCCTGCTGCAGAAACAAIIGAAAT 618
 QY 161 GluSerGlyValArgGluLeuLeuLysArgIleValGlnLysGluAsnThrPheSerAla 180
 DB 619 GAATCAGGTGAAGAGAGCTACTAAAGAGATTGTGCAAGAAAGAAACTGTTCTCTCA 678
 QY 181 PheLeuAsnValLeuArgGlnThrGlyAsnAsnGluValGlnGluLeuThrLysSer 200
 DB 679 TTTCGAATGTTCTTCCTGCAAAACAGAAACAAATGTAATCTGCCAGAGTTAAACAGCT 738
 QY 201 AspCysSerGluSerAsnAlaGluIleGluAsnLeuSerGlnValAspGlyPheGlnVal 220
 DB 739 GATTCCTCAGAAAGCAATGCAGCATGATGAAATTATTCACAAAGTGAATGCTCTCAAC 798
 QY 221 GluGluGlnLeuLeuSerThrThrValGlnProAsnLeuGluLysGluValTyrPheMet 240
 DB 799 GAGAGCAACTCTTTCACACCACTTCAACCAATTCGAGAAAGAGAGCTGTGGGCAATG 858
 QY 241 GluAsnAsnSerSerGluSerPheAlaAspSerSerValValSer 256
 DB 859 GAGATTAACCTCATCAGAAATCATCTTTCACAGATCTTCTGTAGTTTCA 906
 DB
 RESULT 12
 AAS01151
 ID AAS01151 standard; DNA; 1270 BP.
 XX
 AC AAS01151:
 XX
 DT 12-SEP-2001 (first entry)
 XX
 DE Interferon induced nucleic acid, IFN6.
 XX
 DE Interferon induced nucleic acid; autoimmune disease; lupus erythematosus;
 KW immunodeficiency; stroke; acquired immunodeficiency syndrome; AIDS;
 KW graft rejection; viral infection; hepatitis; aplastic anaemia; cancer;
 KW human immunodeficiency virus; HIV; immune-mediated glomerulonephritis;
 KW haematologic disease; chronic neutropenia; myocardial infarction;
 KW neurological disease; Alzheimer's disease; Parkinson's disease; tumour;
 KW amyotrophic lateral sclerosis; spinal muscular atrophy; human; IFN6; ds.
 XX
 OS Homo sapiens.
 XX
 Key Location/Qualifiers
 FH 1..927
 FT CDS
 FT /tag= a
 FT /product= "IFN6"
 FT /note= "Interferon induced polypeptide"
 PN W0200118208-A2.
 PD 15 MAR 2001.
 XX
 PD 08-SEP-2000; 2000MO-US24704.
 XX
 PR 08-SEP-1999; 99US-0152921.
 PR 20-OCT-1999; 99US-0160575.
 PR 20-JAN-2000; 2000US-0177104.
 PR 07-SEP-2000; 2000US-0656633.

PA (CURA-) CUBAGEN CORP.
 PA (BIO) BIOGEN INC.
 XX
 PT Peyman JA, Da Silva A, Hochman P, Hsu A;
 DR MPI: 2001-235201/24.
 XX P-PDB: AAU00297.
 PT New interferon induced polypeptides and polynucleotides useful for the
 PT diagnosis, prevention and treatment of immunological, cell
 PT proliferative disorders, such as lupus erythematosus, cancer, stroke
 PT and Alzheimer's disease
 XX
 PS Claim 9; Page 33-39; 134pp; English.
 XX
 CC The sequence represents interferon induced nucleic acid, IFN5, IFN
 CC nucleic acids and polypeptides are useful for treating or preventing a
 CC pathology associated with IFN polypeptide in a human. They are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of IFN polypeptide or polynucleotide. IFN nucleic
 CC acids, polypeptides and antibodies are useful for diagnosis, prevention
 CC or treatment of variety of immunological and cell proliferative
 CC disorders, such as autoimmune diseases e.g. lupus erythematosus,
 CC immunodeficiency diseases such as acquired immunodeficiency syndrome
 CC (AIDS), graft rejection, viral infections including hepatitis and human
 CC immunodeficiency virus (HIV), immune-mediated glomerulonephritis,
 CC hematologic diseases such as aplastic anemia and chronic neutropenia
 CC and cancer. In addition they are also useful for treating or diagnosing
 CC various disorders associated with cell death, including myocardial
 CC infarction, stroke, neurological diseases including Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis and spinal muscular
 CC atrophy. IFN nucleic acids and polypeptides are also useful for
 CC identifying interferon-like proteins and interferon agonists, for
 CC screening drugs and compounds which inhibit or enhance IFN activity or
 CC function and as targets for the identification of small molecules that
 CC are immunostimulatory, immunosuppressive, or stimulate or suppress normal
 CC cell or tumour cell growth in mammals, including humans.
 XX
 SO Sequence 3692 BP; 1166 A; 709 C; 806 G; 1011 T; 0 other:
 Alignment Scores:
 Pred. No.: 5,32e-70 Length: 3692
 Score: 969.50 Matches: 196
 Percent Similarity: 91.67% Conservative: 2
 Best Local Similarity: 90.74% Mismatches: 5
 Query Match: 18.25% Indels: 14
 DB: 22 Gaps: 1
 US-09-515-363c-2 (1-1025) x AAS01150 (1-3692)
 QY 1 MetSerAsnGlyTyrSerThrAspGluAsnPhenArgTyrLeuIleSerC.sphenArgAla 20
 DB 179 ATGCGAATGGGTATTCACAGAGAGAAATTCGGATCTCATCTCGTATTACGGCC 238
 QY 21 ArgValIysMetTyrIleGlnValAlaGluProValLeuAspTyrLeuThrPrIleuProAla 40
 DB 239 AGGGTGAATAATGTCATCCA----- 258
 QY 41 GluValIysGluGlnIleGlnArgThrValAlaThrSerGlyAsnMetGlnAlaValGlu 60
 DB 259 GAGGTGAAGGAGGATTCAGAGACAGTCCACCTCCGGAAACATGACACGAGTTGAA 318
 QY 61 LeuLeuLeuSerThrLeuGluIuLysGlyValITrphIleSlenGlyTyrTrpIleArgGluPheVal 80
 DB 319 CTGCTGCTGAGCACCTTGAGAGAGGGAGTGGACCTTGCTGGAGCTCCAGAAATTCGTG 378
 QY 81 GluAlaLeuArgThrGlySerProLeuAlaAlaArgTyrMetAsnProGluLeuThr 100
 DB 379 GAGGCCCTCCGAGAAACGGAGCCCTTGCGCCGCCGCTACATGAAACCTTGAGCTTACG 438
 QY 101 AspLeuProSerProSerPheGluAsnAlaHisAspGluTyrLeuGlnIlePheLeuAsnLeu 120
 DB 439 GACTTGCCCTCTCATCTGTTGAGAAACGTCATGATGATATATCTCCAACTGCTGAAACTTC 498

QY 121 LeuGlnProThrLeuValAspIysLeuLeuValArgAspValIleuAspIysCysArgGlu 140
 DB 499 CTTCAGCCCACTCTGTGTGACAAAGCTTCTAGTACAGAGCTTTGGATTAAGTCATGAGAG 558
 QY 141 GluGluLeuLeuThrIleGluAspArgAsnArgIleAlaAlaAlaGluAsnAsnGlyAsn 160
 DB 559 GAGGAACCTGTACCAATTGAACAGCAAAACCGGATTTCGCTGCGAGAAACAAATGCAAT 618
 QY 161 GluSerGlyValArgGluLeuLeuLysArgIleValGlnLysGluAsnTrpPheSerAla 180
 DB 619 GAATCAGGTGTAAAGAGAGCTACTAAAGAGATTGTGCAGAAAGAAAGAACTGCTCTGTGCA 678
 QY 181 PheLeuAsnValLeuArgIleThrGlyAsnAsnGluLeuValGlnGluLeuThrIleLysSer 200
 DB 679 TTCTGAATGTTCTTGTCTCAACAGCAACCAATGCAACTGTCCAAAGCTTAACAGACTCT 738
 QY 201 AspCysSerGluSerAsnAlaGluIleGluAsnLeuSerGlnValAsp 210
 DB 739 GATTGCTGAAAGCAATGCAAGTATTTGTAATTTTACTGAGAGAGAT 780

Search completed: May 16, 2003, 17:37:09
 Job time : 432 secs

GenCore version 5.1 i.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: May 16, 2003, 16:36:02 ; Search time 4440 seconds
(without alignments)
6658.566 Million cell updates/sec

Title: US-09-515-363C-2
Perfect score: 5311
Sequence: 1 MSNGYSTDENFRYLISCPR.....LPITPPNUDSYECCLFSDSD 1025

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_plus_p2n.model -DEV=xlh
-O=/gen2.1/USPTO/spool/US09515363/runat_16052003_100246_24416/app_query.fasta.1.1223
-DB=GenEmbl -QFRT=lastap -SUFFIX=rge -MINMATCH=0.1 -LOOFCUT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human+cdi -LIST=45
-DOALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=5 -MODE=LOCAL
-OUTFMT=fto -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09515363 -CGEN_1_1_2966 -runat_16052003_100246_24416 -NCPU=6 -ICPU=3
-NO_XLPRX -NO_MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGCUT -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELop=6 -DELExt=7

Database : GenEmbl:*
1: gb_da:*
2: gb_hg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_ov:*
22: em_or:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vl:*
30: em_htg_hum:*
31: em_htg_iny:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5311	100.0	3380	9 AF095844	AF095844 Homo sapi
2	5285	99.5	3372	6 AX300832	AX300832 Sequence
3	5285	99.5	3373	9 AY017378	AY017378 Homo sapi
4	4199	79.1	3078	10 AY075132	AY075132 Homo sapi
5	4194	79.0	3771	10 AF374384	AF374384 Mus muscu
6	2874	54.1	1776	9 AK056293	AK056293 Homo sapi
7	2690.5	50.7	2468	10 BC025508	BC025508 Mus muscu
8	2442	46.0	1443	6 AX300841	AX300841 Sequence
9	2271.5	42.8	2050	10 BC004031	BC004031 Mus muscu
10	2021	38.1	1213	9 BC007966	BC007966 Homo sapi
11	1995	37.6	1284	6 AX300838	AX300838 Sequence
12	1356	25.5	2326	10 AF316999	AF316999 Mus muscu
13	1356	25.5	2336	10 BC029209	BC029209 Mus muscu
14	1352	25.5	2584	9 BC019449	BC019449 Homo sapi
15	1344	25.3	2613	9 AK021416	AK021416 Homo sapi
16	1201.5	22.6	1258	6 AX098232	AX098232 Sequence
17	1201.5	22.6	1270	6 AX098236	AX098236 Sequence
18	1129	21.3	2250	9 AK097669	AK097669 Homo sapi
19	1090.5	20.5	3065	9 AF038963	AF038963 Homo sapi
20	1023	19.3	163681	9 AC007750	AC007750 Homo sapi
21	1019.5	19.2	4024	4 AF181119	AF181119 Sus scrofa
22	969.5	18.3	3692	6 AX098234	AX098234 Sequence
23	969.5	18.3	3704	6 AX098238	AX098238 Sequence
24	783	14.7	95417	9 AC010876	AC010876 Homo sapi
25	769.5	14.5	210317	2 AC115074	AC115074 Mus muscu
26	708	13.3	3026	9 AK023661	AK023661 Homo sapi
27	687	12.9	3114	3 AF480439	AF480439 Caenorhab
28	678	12.8	923	10 BC019605	BC019605 Mus muscu
29	642	12.1	392	6 AX071769	AX071769 Sequence
30	636.5	12.0	182906	2 AC111919	AC111919 Rattus no
31	630	11.9	387	6 AX071783	AX071783 Sequence
32	616	11.6	3099	3 AF480440	AF480440 Caenorhab
33	532.5	10.0	118891	2 AC094719	AC094719 Rattus no
34	516	9.7	2633	9 HSM802366	HSM802366 Pyrococu
35	479	9.0	45931	3 U58757	U58757 Caenorhabd
36	479	9.0	226505	2 AC006899	AC006899 Caenorhab
37	472	8.9	301	6 I50897	I50897 Sequence
38	467	8.8	118891	2 AC094719	AC094719 Rattus no
39	457.5	8.6	182906	2 AC111919	AC111919 Rattus no
40	453	8.5	307150	1 CNSPAX01	AJ248283 Pyrococu
41	453	8.5	349980	6 AX041106	AX041106 Sequence
42	441.5	8.3	253505	1 AP000007	AP000007 Pyrococu
43	437	8.2	377	6 A74554	A74554 Sequence
44	437	8.2	377	6 A77533	A77533 Sequence
45	417	7.9	10782	1 AE001002	AE001002 Archaeogl

RESULT 1

ALIGNMENTS

AF095844
LOCUS AF095844 3380 bp mRNA linear Fri 23-JAN-2002
DEFINITION Homo sapiens melanoma differentiation associated protein-5 (MDA5)
ACCESSION AF095844
VERSION AF095844.1 GI:11344593
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Granulata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
AUTHORS Kang, D.-C., Gopalakrishnan, R.V., Wu, Q., Jankowsky, T., Pyle, A.M. and
Fisher, P.B.
TITLE mda-5: An interferon-inducible putative RNA helicase with
double-stranded RNA-dependent ATPase activity and melanoma
growth-suppressive properties
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (2), 637-642 (2002)
MEDLINE 21664412
PUBMED 11805321
REFERENCES 2 (bases 1 to 3380)
AUTHORS Kang, D.-C. and Fisher, P.B.
TITLE Direct Submision
JOURNAL Submitted (30-SEP-1998) Urology, Columbia University, P&S, 630
W168th, New York, NY 10032, USA
FEATURES
source Location/Qualifiers
1..3380
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="HO-1"
/tissue_type="melanoma"
1..3380
/gene="MDA5"
159..3246
/gene="MDA5"
/codon_start=1
/product="melanoma differentiation associated protein-5"
/protein_id="AAG3368.1"
/db_xref="GI:11344594"
/translation="MSNGYSTDENRRLVLSCEPRKVMYQVEPVLDYITPLPAEVE
OIORVATSGNMAVEILLSTLEKGVMLGMRREVEAIKRTGSLARVKNPELID
PSPFENAHDELDLLOPLTDKILVRODVKCMRELLTIEDNRVIAANNNG
ESGVRELLKRVOKENMSATLVNIRQGNNEIKVQELTNSGSSNAIENLSDYD
OVEDLLSTTVQPNLEKEVGMENSSFSRSDSVSTSLAEVSGSLDLSLGH
NSNMGSDGTMGSDSDENVARASPEPELQIKPQMVAPALLEGKNTIICLTGSG
KTRVAVYIAKHDLKKKASEPGKAVIYANVLLVEQLKKEQPLKKMYVGLSG
DTQLKISPEVYKSCDIIITQIILENLINLENEIDAQVOLDPSLIIECHHTNK
EAVYNNIRHRLIMOKLKNRLKKEKKPVIPDQILIGLASGGATGQAAREHILK
LCANDARTITVYENLDLKNLODERKKAIDATIRPKKLEIETMRITQYCO
MSPHSDFTQPTQONAIOMENKAAKGRKRVCAEHLKTNELQINDTIRMTDAT
HLETTYNEEKQKRAVIEDSDGDDCYCQIDDEDQIKKPLKIDETRIIMLTPE
NNKMLKLAENPEYENKLTAKRLNTIMQYRITESASJLETKTQSAVYLSQMTLE
NKEFAEVCAHHHLGAGHSEPKPTONBOKEPVISKPRCKINLLATTVAEGLDI
KRCNVIVRYGLTNEIAMVQARGARARAESEYVIVAAHSGVITETTYNDREKMYK
AIRCYQNNKPREPAKILIELOMOSIMEKKTKKIAKTKNNSTLTETUCKNSVLA
CSGEDIHVIEKHMHVNMTPERKELYVENKALCKKAMVQINDEIICKQGAAGTMM
VHKGLDLCLEIRNFVYVYKNNSTKKQKWELEPITPHNIDYSECLFSDP"

BASE COUNT 1153 a 644 c 753 g 830 t
ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 3380
Score: 5311.00 Matches: 1025
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-515-363c-2 (1-1025) x AF095844 (1-3380)

OY 1 MetSerAsnGlyTyrSerThrAspGluAsnPheArgTyrLeuIleSerTyrPheArgVala 20
DB 169 ATGTGCAATGGGATATTCACAGACGAGAAATTCCTCGCTATCATCTCGTCAAGGAGCC 228

OY 21 ArgValIysMetTyrIleGlnValGluProValLeuAspTyrLeuThrPheLeuIleVala 40
DB 229 AGGTGAAATATGTACATCATCAGATGAGCCTGTGCTGACACTACCTGACCTTCTCTCTCA 288
OY 41 GluValIysGluGlnIleGlnIleArgThrValAlaThrSerGlyAsnMetGlnAlaVala 60
DB 289 GAGGTGAAGAGACAGATTCAGAGGACAGCTGCCACCTCGGGAACATGACAGCAATGAA 348
OY 61 LeuLeuLeuSerThrLeuGlnIleGlyValIlePheIleGlyTyrPheArgGlnVala 80
DB 349 CTGCTGCTGACGACCTTGGAGAGAGGAGTCTGGCACCTTGGTGGACCTGGGAAATCTGTG 408
OY 81 GluAlaLeuArgArgThrGlySerProLeuAlaAlaArgTyrMetAsnProGlnThr 100
DB 409 GAGGCGCTCCGAGAGAACCGGACGCGCTCGGCGCGCGCTCATAGAACCTGTGACG 468
OY 101 AspleuProSerProSerPheGluAsnAlaIleAspGluTyrLeuGlnLeuLeuVala 120
DB 469 GACTTGCCCTCTCCATCTGCTTGGAGAACCTCATGTGAAATATCTCCAACTGCTCTATCTG 528
OY 121 LeuGlnProThrLeuValAspLysLeuLeuValArgAspValLeuAspLysCysVala 140
DB 529 CTTGACGCCACTCTGTGTGAGACAGCTTCACTTAGACGCTTGGATTAAGTCTCTAGAG 588
OY 141 GluGluLeuLeuThrIleGluAspArgAsnArgIleAlaAlaGluAsnAsnVala 160
DB 589 GAGGAACCTGTGACAAATTTGAAGACAGAAACCGGATTCCTGCTGACAGAAACAAAT 648
OY 161 GluSerGlyValArgGluLeuLeuLysArgIleValGlnLysGluAsnTrpVala 180
DB 649 GAATGAGGTGTAAAGACACTACTATAAAAGATTGTGCGAAAGAAAGAAATGCTGTCTTCA 708
OY 181 PheLeuAsnValLeuArgGlnThrGlyAsnAsnGluLeuValGlnGluLeuVala 200
DB 709 TTTCGAATGTTCTTCTGCTCAACAGAAACAAATCACTTGTCCAAAGATTAAATCTCT 768
OY 201 AspCysSerGluSerAsnAlaGluIleGluAsnLeuSerGlnValAspGlyProVala 220
DB 769 GATTCCTCAGAAAGCAATGACAGATTCAGAAATTTATCAGAAAGTGAAGTGCCTCAAGT 828
OY 221 GluGluGlnLeuLeuSerThrThrValGlnProAsnLeuGluLysGluValaIleVala 240
DB 829 GAAAGCAACTCTTCTTCAACCAAGATTCAGCAATTCGAGAAAGAGGTGTGCAATG 888
OY 241 GluAsnAsnSerSerGluSerSerPheAlaAspSerValValSerGluSerVala 260
DB 889 GAGATTAATCTCATGAAATCATCTTTTCCAGATTCTTCTGTAGTTTGAGATTCATACA 948
OY 261 SerLeuAlaGluGlySerValSerCysLeuAspGluSerLeuGlyIleAsnSerVala 280
DB 949 AGTTGGCAGAAAGCAAGTGCAGCTGCTTGAATCAAACTCTTGGACATTAACACGACATG 1008
OY 281 GlySerAspSerGlyThrMetGlySerAspSerAspGluGluAsnValAlaAlaVala 300
DB 1009 GGCAATGATTCAGGACCATGAGGAGTATTCACATGAAGAAATGTGGCAATGAAGCA 1068
OY 301 SerProGluProGluLeuGlnLeuArgProTyrGlnMetGluValAlaGlnIleVala 320
DB 1069 TCCCGGAGCCAGAACTCAGCTCAGGCTTACCAAAATGGAAGTTGCCAGACGACGACCT 1128
OY 321 GluGlyLysAsnIleIleIleCysLeuProThrGlySerGlyLysThrArgValAlaVala 340
DB 1129 GAAGGAAATATATCATCATCTGCTCCCTACAGGAGTGGAAAAACACAGACTCTGTGT 1188
OY 341 TyrIleAlaLysAspIleLeuAspLysLysLysAlaSerGluProGlyIleVala 360
DB 1189 TACATTGCCAAGATCACTTGAACAGAAAGAAAGATTCGACCTCGCAAAATTTATA 1248
OY 361 ValLeuValAsnLysValLeuLeuValGlnGlnLeuPheArgLysGluPheGlnIleVala 380
DB 1249 GTTCTTGCATTAAGGTACTGCTAGTTGACACAGCTCTTCCGCAAGGAGTCTTAATTT 1308

QY 381 LeuLysLysTrpTyrArgValIleGlyLeuSerGlyAspThrGlnLeuLysIleSerPhe 400
 DB 1309 TTGAGAAATGGATGCTGTTATTGGATTAAAGGTGATACCACTGAAATAATATCATTT 1368
 QY 401 ProGluValValLysSerCysAspIleIleIleSerThrAlaGlnIleLeuGluAsnSer 420
 DB 1369 CCGAAGATTGTCAAGTCCTGTGATTATATATCATGACAGCTCAATCCCTGAAACTTC 1428
 QY 421 LeuLeuAsnLeuGluAsnGlyAspAlaGlyValGlnLeuSerAspIleSerLeuIle 440
 DB 1429 CTCTTAACTTGAATAATGAGAAAGATGCTGCTGTTCAATGTCACAACTTTCCTCAT 1488
 QY 441 IleIleAspGluCysHisHisThrAsnLysGluAlaVal-TyrAsnAsnIleMetCysHis 460
 DB 1489 ATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1548
 QY 461 TyrLeuMetGlnLysLeuLysAsnAsnArgLeuLysLysGluAsnLysProValIlePro 480
 DB 1549 TATTGATGCGAAGTTGAAACCAATAGACTCAAGAAAATAAACCAACTGATGATCC 1608
 QY 481 LeuProGlnIleLeuGlyLeuThrAlaSerProGlyValGlyGlyAlaThrLysGlnAla 500
 DB 1609 CTTCCTCAGATCTGGACATACAGCTTACCTGCTGTGGAGGCGCAAGACAGCC 1668
 QY 501 LysAlaGluGluHisIleLeuLysLeuCysAlaAsnLeuAspAlaPheThrIleLysThr 520
 DB 1669 AAGGCTCAACACACATTTTAAACATATGCGCAATCTGATCATTTTATTTAAACT 1728
 QY 521 ValLysGluAsnLeuAspGlnLeuLysAsnGlnIleGlnGluProCysLysPheAla 540
 DB 1729 GTTAAACAAACCTTCACTCACTGAAACCAATACAGCAACCACTGCAAGAGTTG 1788
 QY 541 IleAlaAspAlaThrArgGluAspProPheLysGluLysLeuGlnIleMetThrArg 560
 DB 1789 ATTGCAATGCAACACAGACAGATCCATTTAAAGAGAACTTCTAGAAATAAGACAA 1848
 QY 561 IleGlnThrTyrCysGlnMetSerPrometSerAspPheGlyThrGlnProTyrGluAla 580
 DB 1849 ATTCAAACTTATTGTCAAATGAGTCCAAATGTCAGATTGTCAACTCACTGTAACAA 1908
 QY 581 TrpAlaIleGlnMetGluLysLysAlaAlaLysLysGlyAsnArgGlyGluArgValCys 600
 DB 1909 TGGGCCATTCATGAAAAAGAAAAAGCTGCAAAAAAGAAATGCAAAAGACGCTGTT 1968
 QY 601 AlaGlnHisLeuArgLysTyrAsnGluAlaLeuGlnIleLeuAsnProThrIleArgMetIle 620
 DB 1969 GCGAACAATTTGAGAGAGTACAAATGAGGCCCTACCAATTAATGACACAAATGCA 2028
 QY 621 AspAlaThrHisLeuGluThrPheTyrAsnGluGluLysAspLysIlePheAlaVal 640
 DB 2029 GATGCGTATCTCATCTTGAACCTTTCTATTAATGAAAGACAAATGAAGAGTTGCA 2088
 QY 641 IleGluAspSerAspArgGluGlyLysAspAspGluTyrCysAspGlyAspGlu 660
 DB 2089 ATGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2148
 QY 661 AspAspLeuLysLysProLeuLysLeuAspGluThrAspArgPheLeuMetThrLeuPhe 680
 DB 2149 GATGATTTAAGAAACCTTTGAAACTGGATGAACACAGATGATTTTCAAGCTTAA 2208
 QY 681 PheGluAsnAsnLysMetLeuLysArgLeuAlaGluAsnProGluTyrGluAsnGluLys 700
 DB 2209 TTTGAAACAAATTAATAATGTTGAAAGGCTGCTGAAACCCAAATATGAAATGAAAG 2268
 QY 701 LeuThrLysLeuArgAsnThrIleMetGluGlnIleTyrThrArgThrGluGlnSerAlaArg 720
 DB 2269 CTGACCAAAATTAAGAAATACCAATATGACCAATATACAGACTGAGATGACACCA 2328
 QY 721 GlyIleIlePheThrLysThrArgGlnSerAlaTyrAlaLeuSerGlnTrpIleThrGln 740
 DB 2329 GGAATATACTTTTACAAAAACAGACAGAGATATGCGCTTTCCCAAGTAAATATCTGA 2388
 QY 741 AsnGluLysPheAlaGluValGlyValLysAlaHisHisLeuIleGlyAlaGlyHisSer 760

DB 2389 AATGAAATAATTTGCTGAAGTACAGAGTCAAAAGCCACCAATCTGATTTGAGACTGACACAG 2448
 QY 761 SerGluPheLysProMetThrGlnAsnGluGlnValIleSerLysPheArgThr 780
 DB 2449 AGTGAGTTCAAAACCCATGACAGATGAAACAAAAAGAAAGTCAATTAAGTAAATTCGCA 2508
 QY 781 GlyLysIleAsnLeuLeuIleAlaThrThrValAlaGluGluGlyLeuAspIleLysGlu 800
 DB 2509 GGAATAATCAATGCTTATGCTTACACAGAGTGGCAGAAAGGCTGTGATATTAACA 2568
 QY 801 CysAsnIleValIleArgTyrGlyLeuValThrAsnGluIleAlaMetValGlnAlaArg 820
 DB 2569 TGTAACTGTTGATGCTGATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2628
 QY 821 GlyArgAlaArgAlaAspLysThrTyrValLeuValAlaHisSerGlySerGlyVal 840
 DB 2629 GGTGAGCCAGAGCTGATGAGAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2688
 QY 841 IleGlnHisGluThrValAsnAspPheArgGluLysMetMetTyrLysAlaIleHisCys 860
 DB 2689 ATGCAACATGACACATTTATGATTTCCGAGAGAGATATGATTAAGCTATACATTT 2748
 QY 861 ValGlnAsnMetLysProGluGluTyrAlaHisLysIleLeuGluLeuGlnMetGlnSer 880
 DB 2749 GTTCAAAATATGAAACCCAGAGAGTATGCTCATTAAGATTTTGAATTAACAGAT 2808
 QY 881 IleMetGluLysLysMetLysThrLysArgAsnIleAlaLysHisTyrLysAsnPro 900
 DB 2809 ATATATGAAAGAAATATGAAACCAAGAGAAATATGCTCAACATTAACAAATTA 2868
 QY 901 SerLeuIleThrPheLeuGlyLysAsnGlySerValLeuAlaCysSerGlyLysAla 920
 DB 2869 TCACTAATTAATCTTCTTGTGCAAAAACCTGACGTGCTGCTGCTGCTGCTGCTGCT 2928
 QY 921 HisValIleGluLysMetHisLysValAsnMetThrProGluPheLysGluLeuArg 940
 DB 2929 CATGTAATGAGAAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2988
 QY 941 ValArgGluAsnLysAlaLeuGlnLysLysCysAlaAspTyrGlnIleAsnGlyVal 960
 DB 2989 GTAGAGAAACCAACACCTGCAAAAAAGTGTGCGCATATCAATAATTAATTA 3048
 QY 961 IleCysLysGlyGlnIleArgPheLysMetValHisLysGlyLeuAspLeuPro 980
 DB 3049 ATCTGCAATATGCGAGCTTGGGACACAAATGATGTGTCACAAAGCTTACATTTGCT 3108
 QY 981 CysLeuLysIleArgAsnPheValValPheLysAsnAsnSerThrLysLysIleTyr 1000
 DB 3109 TGTCTCAAAATATAGCAATTTTGTAGTGCTTTTCAAAAATTAATTCACAAAGAACAA 3168
 QY 1001 LysLysTrpValGluLeuProIleThrPheProAsnLeuAspTyrSerGlyCysLysLeu 1020
 DB 3169 AAAAAGTGGTGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3228
 QY 1021 PheSerAspGluAsp 1025
 DB 3229 TTTAGTATGAGGAT 3243
 RESULT 2
 AX300832 3372 bp DNA linear Jan 30-Nov-2001
 LOCUS AX300832
 DEFINITION Sequence 1 from Patent WO0185955.
 ACCESSION AX300832
 VERSION AX300832.1 GI:17382110
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1
 Bahr, G., Cocude, C. and Capron, A.
 TITLE Rh16 polypeptides and its fragments and polynucleotides encoding

Journal Patent: WO 0185935-A 1 15-NOV-2001;
 Istac (FR) : INSTITUT PASTEUR DE LILLE (FR)
 Location/Qualifiers
 1. 3372

FEATURES
 source /organism="Homo sapiens"
 /db_xref="taxon:9606"
 155. 3232
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAD13049.1"
 /db_xref="GI:17382111"

BASE COUNT 1157 a 637 c 748 g 830 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 0 Length: 3372
 Score: 5285.00 Matches: 1020
 Percent Similarity: 99.71% Conservative: 2
 Best Local Similarity: 99.51% Mismatches: 3
 Query Match: 99.51% Indels: 0
 DB: 6 Gaps: 0
 US-09-515-363c-2 (1-1025) x AX300832 (1-3372)

QY 1 MetSerAsnGlyTyrSerThrAspGluAsnProPheArgTyrLeuIleSerGlyPheArgAla 20
 DB 155 ATGTCCAAATGGGATTATCCACAGCGAGAAATTCCTGCTATCTCAATCCGCTATTCCAGCGGC 214
 QY 21 ArgValLysMetTyrIleGlnValGluProValLeuAspTyrLeuThrPheLeuProAla 40
 DB 215 AGGGTGAATAATGTACATCCAGGTGAGCGCTGCTGCTACTTACCTTCTGCTCTGA 274
 QY 41 GluValLysGluGlnIleGlnArgThrValAlaThrSerGlyAsnMetGlnAlaValGln 60
 DB 275 GAGGTGAGAGAGAGATTCAGAGAGACAGTCCGACCTCCGAGAAATATGAGCCAGATTGA 334
 QY 61 LeuLeuLeuSerThrLeuGluLysGlyValIleThrHisLeuGlyTyrThrArgIlePheVal 80
 DB 335 CTGCTCTGCTAGCACTTGGAGAGAGAGTCCGACCTTGGTGGTGAATCTGAGAAATCTG 394
 QY 81 GluAlaLeuArgArgThrGlySerProLeuAlaAlaArgTyrMetAsnProGluLeuThr 100
 DB 395 GAGGCCCTCCGAGAACCCGCGACCTCTGGCGCGCGCTACATGACCTTGAAGCTCAGCG 454
 QY 101 AspLeuProSerProSerPheGluAsnAlaHisAspGluTyrLeuGlnIleLeuAsnLeu 120
 DB 455 GACTTCCTCCCTCATCGTTTGGAGAGCGCTCATGATGATATCTCCAAATTCGTGAACCTC 514
 QY 121 LeuGlnProThrLeuValAspLysLeuLeuValArgAspValLeuAspLysGlyMetGlu 140
 DB 515 CTTCACCCCACTCTGTGTGACAAAGCTTCTAGTAGACCTCTTGCATATATGCAATGAGG 574
 QY 141 GluGluLeuLeuThrIleGluLysArgAsnArgGluLeuAlaAlaGluLysAsnGlyAsn 160

DB 575 GAGGAACCTGTTGACAAATTTGAAGACAGAAACCGGATTCCTGCTGCAAGAAACATTTGA 634
 QY 161 GluSerGlyValArgGluLeuLeuLysArgGluValGlnLysGluAsnTyrPheValAla 180
 DB 635 GAATCAGGTGTAGAGAGACTACTAAAGATTTGTGCAAGAAAGAAAGAACTGGTCTCTGCA 694
 QY 181 PheLeuAsnValLeuArgGlnThrGlyAsnAsnGluLeuValGlnGluLeuThrTyrSer 200
 DB 695 TTTCGAATGTTCTTCGCAACAGAGAAACATGACCTTGCACGACTTACACACTCTCT 754
 QY 201 AspCysSerGlySerAsnAlaGluIleGlnLysLeuSerGlnValAspGluProGluVal 220
 DB 755 GATTGCTCAGAAAGCAATGCAAGATTTAGAAATTTATACAAAGTTGATGCTTCAAGT 814
 QY 221 GluGlnGlnLeuLeuSerThrThrValGlnProAsnLeuGluLysGluValIleTyrMet 240
 DB 815 GAGAGCAAACTCTTCAACACCACTTCAGCCAAATCTGGAAGAGAGCTGGCTGCTATG 874
 QY 241 GluAsnAsnSerSerGluSerSerPheAlaAspSerSerValValSerGluSerAspThr 260
 DB 875 GGAATTAATCATCATGAAATCATCTTTGCAATTCCTGAGATTGCAAAATCAACATCA 934
 QY 261 SerLeuAlaGluGlySerValSerCysLeuAspGluSerLeuGlyHisAsnSerAsnMet 280
 DB 935 ACTTTGGCAGAGAGAAAGTCTAGCTGCTTACGTGAAAGCTTTGGACATACACAGATATG 994
 QY 281 GlySerAspSerGlyThrMetGlySerAspSerAspGluGluAsnValAlaAlaValAla 300
 DB 995 GCGAGTGAATTCAGGACCATGGAAGTGTACATGAAAGAAATGTGCGACAGCAAGACCA 1054
 QY 301 SerProGluProGluLeuGlnLeuArgProTyrGlnMetGluValAlaGlnProValLeu 320
 DB 1055 TCCTCCGAGACCGAAGCTCAGCTCAGGCTTACCAAAAGGAAGTTGCCACCACTTGTG 1114
 QY 321 GluGlyLysAsnIleIleIleCysLeuProThrGlySerGlyLysThrArgValAlaVal 340
 DB 1115 GAGGAGAAATATATATCATCTGCTGCTCCTACAGGAGATGGAAGAAACAGAGTGTGTT 1174
 QY 341 TyrIleAlaLysAspHisLeuAspLysLysLysLysAlaSerGluProGlyLysValIle 360
 DB 1175 TACATTGCCAACAGATCACTTACAGCAAGAAAGAAAGCATCTGACCTCGGAAGATATA 1234
 QY 361 ValLeuValAsnLysValLeuLeuValGluGlnLeuPheArgLysGluPheGlnProThr 380
 DB 1235 GTTCTTGCAATTAAGCTAGTGTAGTGAACAGCTCTTCCGAGAGAGTTCACACATTT 1294
 QY 381 LeuLysLysTyrTyrArgValIleGlyLeuSerGlyAspThrGlnLeuLysIleSerThr 400
 DB 1295 TTGAAGAAATGATATCGATTATGATTAAGTGTGATACCAACATGCAAAATATCATTT 1354
 QY 401 ProGluValValLysSerCysAspIleIleIleSerThrAlaGlnIleLeuGluAsnSer 420
 DB 1355 CCAGAAAGTTGCAAGTCTGTGATATTATTCATACATGCTCAAAATCTTGAAGAACTCC 1414
 QY 421 LeuLeuAsnLeuGluAsnGlyGluAspAlaGlyValGlnLeuSerAspPheSerLeuIle 440
 DB 1415 CTTTAAACTTGGAAAGAGAGAGATCTGCTGTTCAATTGTCAAGCTTTCCTTCATTT 1474
 QY 441 IleIleAspGluCysHisIleIleThrAsnLysGluAlaValTyrAsnAsnIleMetArgHis 460
 DB 1475 ATCATTTGATGATATGATCATCACACCAAGAGAGAGTATTAATACATCATGAGGATAT 1534
 QY 461 TyrLeuMetGlnLysLeuLysAsnAsnArgLeuLysLysGluAsnLysProValIlePro 480
 DB 1535 TATTGATTCACAGAACTTGAAGAAATATGACATCAAGAAAGAAACCAACAGATATCCC 1594
 QY 481 LeuProGlnIleLeuGlyLeuThrAlaSerProGlyValGlyGlyAlaThrLysGlnAla 500
 DB 1595 CTTCCTCAGATACCTGGAGCTAACACCTTCACTGCTGTGGAGGCGCCACAGCAACCTC 1654
 QY 501 LysAlaGluGlnHisIleLeuLysLysCysAlaAsnLeuAspAlaPheThrIleLysThr 520
 DB 1655 AAGCTGAAGAACACATTTTAAACTATGTGCCAATCTTGATGCTATTACTATTAAATTT 1714

QY 521 ValLysGluAsnLeuAspGlnLeuLysAsnGlnIleGlnGluProCysLysPheAla 540
 DB 1715 GTTAAAGAAAACCTGATCACTAGTAAAAACCAATATACAGAGCTATGCAGAAAGTTGGC 1774
 QY 541 IleAlaAspAlaThrArgGluAspProPheLysGluLysLeuGluIleMetThrArg 560
 DB 1775 ATTGCAGATGCAACGACAGACAGATCCATTAAAGAAACCTTCTAGAAATATGACAAAG 1834
 QY 561 IleGlnThrTyrCysGlnMetSerProMetSerAspPheLysPheGlnProTyrGluGln 580
 DB 1835 ATTCAAACTATTGCAATGAGTCCCAATGAGATTGAGACCTCAACCTATATACAA 1894
 QY 581 TPAlaIleGlnMetGluLysLysAlaAlaLysGluLysAsnArgLysLysValGly 600
 DB 1895 TGGGCCATTCAATGAGAAAAAAGCTGCAGAAAGCAATACCAAGAAAGTGTTCCT 1954
 QY 601 AlaGluHisLeuArgLysTyrAsnGluAlaLeuGlnIleAsnAspThrIleArgMetIle 620
 DB 1955 GCAGAACATTGAGAGACATACATTAAGCCCTACCAATATATACCAATACGAAAGATA 2014
 QY 621 AspAlaTyrThrHisLeuGluThrPheTyrAsnGluLysLysPheAlaVal 640
 DB 2015 GATGGCTATACCTCATCTGAAACTTCTATTAATGAAAGCAAAAGATAGAGATTGGCAGTC 2074
 QY 641 IleGluAspAspSerAspGluGluGluAspAspGluTyrCysAspGluAspGlu 660
 DB 2075 ATAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2134
 QY 661 AspAspLeuLysLysProLeuLysLeuAspGluThrAspArgPheLeuMetThrLeuPhe 680
 DB 2135 GATGATTAAAGAAACCTTGAAGAGTGAAGAGATGAGATTTCTCATGACTTATTT 2194
 QY 681 PheGluAsnAsnLysMetLeuLysArgLeuAlaGluAsnProTyrTyrGlnAsnGluLys 700
 DB 2195 TTGGAACAATTAATGTTGAAAAGGCTGGCTGAAAACCCAGAAATATGAAATGAAAG 2254
 QY 701 LeuThrLysLeuArgAsnThrIleMetGluGlnTyrThrArgPheGluLysLeuAlaArg 720
 DB 2255 CTGACCAAAATTAAGAAATGCAATGAGCAATATACATGAGATGAGAGAAACAGACCA 2314
 QY 721 GlyIleIlePheThrLysThrArgGlnSerAlaTyrAlaLeuSerGlnProIlePheArg 740
 DB 2315 GGAATATCTTTACAAAACAGACAGAGATGCATATGCCCTTCCAGAGAAATTATGAA 2374
 QY 741 AsnGluLysPheAlaGluValGlyValLysAlaHisLysLeuIleGlyAlaGlyHisSer 760
 DB 2375 AATGAAAAATTTGCTGAACTAGAGTCAAGCAAGCCACCATTTGATGAGAGTGCAGACAGC 2434
 QY 761 SerGluPheLysProMetThrGlnAsnGluGlnLysGluValIleSerLysPheArgTyr 780
 DB 2435 AGTGAGTTCAAAACCCATGACAGCAATGAAACAAAGAAATCAATGAAATTTCCACAT 2494
 QY 781 GlyLysIleAsnLeuLeuIleAlaThrValAlaGluGluLysLeuAspIleCysGlu 800
 DB 2495 GGAATAATTAATCTGCTTATGCTACAGAGTGGCAGAGAGAGTGTGATATAAGAA 2554
 QY 801 CysAsnIleValIleArgTyrGlyLeuValPheAsnGluIleAlaMetValIleAlaArg 820
 DB 2555 TGTAAATGTTATGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2614
 QY 821 GlyArgAlaArgAlaAspGluSerThrTyrValIleValAlaHisSerGlySerGlyVal 840
 DB 2615 GCTCGAGCCAGAGCTGATAGACACCTAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2674
 QY 841 IleGlnHisGluThrValAsnAspPheArgGluLysMetMetLysLysAlaIleHisCys 860
 DB 2675 ATGGAAGCGAGACACTATATGATTTCCGAGAGAAATGATGATATTAAGTATATAGCTTCT 2734
 QY 861 ValGluAsnMetLysProGluGluTyrAlaHisLysIleLeuGluLeuLeuGlnSer 880
 DB 2735 GTTCAAAATATGAAACCAAGAGAGTATGCTCATTAAGATTTGGAATTAACAGATGCAAGT 2794

QY 881 IleMetGluLysMetLysThrLysArgAsnIleAlaLysHisTyrLysAsnAsnPro 900
 DB 2795 ATTAATGGAAGAAATTAATAAACAGAGAAATATTTGCCAGAGATTACAGAAATATACCA 2854
 QY 901 SerLeuIleThrPheLeuCysLysAsnCysSerValLeuAlaCysSerGlyLysAspIle 920
 DB 2855 TCACATTAATACCTTCCCTTGCAGAAAACCTGAGTGTGAGCTGTTCCTGCGGAAATATC 2914
 QY 921 HisValIleGluLysMetHisHisValAsnMetThrProGluPheLysGluLeuPyrIle 940
 DB 2915 CATGTAATGGAAGAAATCAATCAGTCATATGAGCCGAGAAATTCAGGAATTTTATAT 2974
 QY 941 ValArgGluAsnLysAlaLeuGlnLysLysCysAlaAspTyrGlnIleAsnGlyGluIle 960
 DB 2975 GTAAGAGAAACCAAGCACTGCAAAAGAGTGTGCCAGCTATCAATTAATATGCGAAATC 3034
 QY 961 IleCysLysCysGlyGlnAlaTyrPheLysThrMetValHisLysGlyLeuAsnLeuPro 980
 DB 3035 ATCTGCAATGTGGCCAGCTTGGGGAACATGATGTGCAACAAAGGCTTAAATTTGCT 3094
 QY 981 CysLeuLysIleArgAsnPheValValAlaPheLysAsnAsnSerThrLysLysGlnTyr 1000
 DB 3095 TGTCTCAAAATAAGGAATTTGTAGTGTTCCTCAAAATAATTCAGCAAGAAACATATC 3154
 QY 1001 LysLysTyrValGluLeuProIleThrPheProAsnLeuAspTyrSerGluLysCysLeu 1020
 DB 3155 AAAAGCTGGTGAATTAACCTATATCCATTTCCCAATCTTGACTATTCAGAAATGCTTTA 3214
 QY 1021 PheSerAspGluAsp 1025
 DB 3215 TTTAGTATGAGAGAT 3229
 RESULT 3
 AYO17378 3373 bp mRNA linear HKI 31-JAN-2001
 LOCUS AYO17378
 DEFINITION Homo sapiens RNA helicase-DEAD box protein RH116 mRNA, complete cds.
 ACCESSION AYO17378
 VERSION AYO17378.1 GI:12621065
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 3373)
 Capron, A., and Bahr, G.M.
 Cocude, C., Kolesnitchenko, V., Billaut-Mulot, O., Truong, M.-J., Immunomodulator Murabutide Identification of a new RNA helicase (RH116) regulated by the unpublished 2 (bases 1 to 3373)
 JOURNAL 2 (bases 1 to 3373)
 REFERENCE Cocude, C., Kolesnitchenko, V., Billaut-Mulot, O., Truong, M.-J., Capron, A. and Bahr, G.M.
 TITLE Direct Submission
 JOURNAL Submitted (09-JAN-2001) Laboratoire d'Immunologie Moléculaire de l'Infection et de l'Inflammation, Institut Pasteur de Lille, 1 rue du Professeur Calmette BP 245, Lille 59019, France
 FEATURES
 source
 1..3373
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 155..3232
 /codon_start=1
 /product="RNA helicase-DEAD box protein RH116"
 /protein_id="AA054076.1"
 /db_xref="GI:12621066"
 /translation="MSNGSYSDENFRFLISCFRAKVKYIOVEPVLLYTFLEPAEYVE
 OIOEVATSGNMAVEILLISTLEGWHLHETREFVALRRTGSLAARYMNPETLD
 PPSFENAHDETDLLNLOFTLDKLLVRYDVKCEBELITIEDNRRLAAABNNON
 ESEVRELKRIYQENMFSAFLNLRGTNNELVQELTGDSSNAEINLSQVDDP
 QVEGOLLSTYVQPNLEKENVGAMENSSSEFPAQSVESDTSIAEGSVGLDESLGH
 NSNGSPSGTGSDDENVAARASPEPELDRLRYQEVNAPALIEGNIILICPTGSG
 KTRAVAYIAKHLDKKKKSEPKVIVLYNKKVLLVEQLPFKEFPLKKMYRVIIGLSG

QY	1	MetSerAangLYTYrSerThrAspLysAsnProMetArgTYrLeuLeuSerCysPheArgVal	20
Db	155	ATGTGAAATGGGATATCCACAGACAGAAATTTCCCTATTCATCTGCTGCTTCAGGGCT	214
QY	21	ArgValLysMetTYrIleGlnValGluProValLeuAspTYrIleThrPheLeuProAla	40
Db	215	AGGGTAAAGATGATCATCCAGTGGAGGCCCTGGTGGCACTACCTGACCTTCCTGGCTGTA	274
QY	41	GluValLysGluGlnIleGlnArgThrValAlaThrSerGlyAsnMetGlnAlaValGlu	60
Db	275	GAGGTGAAGAGACAGATTCAGAGACAGCTGGCCACTCTGGGAACATGTAAGGCAATTGCA	334
QY	61	LeuLeuLeuSerThrLeuGluLysGlyValTYrPheIleLeuLYrTYrThrArgLuePheVal	80
Db	335	CTGGCTGGACACACTTGAGAAAGGAGATCTGGCACCTTGCTGGAGCTGGGCATTTGGG	394
QY	81	GluAlaLeuArgArgThrGlySerProMetAlaIleArgTYrMetAsnProGluLeuThr	100
Db	395	GAGGCTCTCCGGAGAACGGAGCCCTCTGGCGGGCCGCTACATGAACTCTGAGTTCAG	454
QY	101	AspLeuProSerProSerPheGluAsnAlaHisAspGluTYrLueGlnLueLeuAsnArg	120
Db	455	GACTTGCCTCTCCATCTGTTGAGAAACCTCATGTGAAATATCTCCAAATCTGTAAACCTG	514
QY	121	LeuGlnProThrLeuValAspLysLeuLeuValArgAspValLueAspLysCysMetGlu	140
Db	515	CTTAGCCCACTCTGGTGGACAAAGCTTAGTTAGTAGAGCGTCCTGGATTAAGTCATGGAG	574
QY	141	GluGluLeuLeuThrIleGluAsnParGAsnArgIleAlaAlaIleAlaGluAsnAsnGlyAsn	160
Db	575	GAGAGACTGTTGACATTTGAACAGCAAGAAACCGGATTCGTCTGTAGAAATCAATGGAAT	634
QY	161	GluSerGlyValArgGluLeuLeuLysArgIleValGlnLysGluAsnTrpPheSerAla	180
Db	635	GAATCAGCTGTAAAGACACTACTAAAAAGGATTGGCGAAAGAAACCTGGTCTCTGCA	694
QY	181	PheLeuAsnValLeuArgGluThrGlyAsnAsnGluLeuValGlnGluLeuThrGlySer	200
Db	695	TTTCTGAAATGTTCTTCGTCAACAGGAAACAAATGAATCTTGCTCAAGACTTAAACAGCTT	754
QY	201	AspCysSerGluSerAsnAlaGluIleGluLysLeuSerGlnValAspTYrProGlnVal	220

Db	755	GATTGGCTACGAAACCATGCGAGATTTGAGAAATTTATCAAGAAAGTTGATGGCTCCAAATG	814
OY	221	GIUGUGIULEULEUSeRtThrValGlnProAsnLeuGluValIlePrlMet	240
Db	815	GAAAGCAACTCTTCTTCACCAACAGTTTCAGCCAAATCTGGACAGAGCTGGGCAATG	874
OY	241	GIUAsnAsnSerSerGluSerSerPheAlaAspSerSerValValSerGluSerAspThr	260
Db	875	GAGAAATACATCATCAGAAATCATCTTTTGGCAGATTTCTTGTAAGTTCCAGAAATCAGAAC	934
OY	261	SeRLeuAlaGluSerValSerCysLeuAspGluSerLeuGluHisAsnSerAspMet	280
Db	925	AGTTGGCAGAGAAAGATGTCACCTGCTAGATGAAGAACTTGGACATTAACAGCAATG	994
OY	281	GlySerAspSerGlyThrMetGlySerAspSerAspGluAsnValAlaAlaValAla	300
Db	995	GGCAGTATTTCAGGCACCAATGGGCAAGTGTATTCAGATGAAGAAATGTGGCAGCAAGCA	1054
OY	301	SeRProGluProGluLeuGluLeuAlaGProTyrGlnMetGluValAlaGluProAlaLeu	320
Db	1055	TCCCCGGAGCCAGACCTCCAGCTCAGGCCCTTACCAATGTGAAGTTGCCAGCACCTTG	1114
OY	321	GIUGIlyAsnIleIleIleCysLeuProThrGlySerGlyLyThrArgValAlaVal	340
Db	1115	GAAAGCAAGATATCATCATCTGCTGCTCCACAGGAGTGGAAAAACCAAGATGCTGTT	1174
OY	341	TyrIleAlaAspHisLeuAspLysLysLysValAlaSerGluProGlyLysValIle	360
Db	1175	TACATTGCCAAGGATCTCACTTAGCAAGAAAAAGCAATCTGACCTGGAAAAATATATA	1234
OY	361	ValLeuValAsnLysValLeuLeuValGluGlnLeuPheArgLysGluPheGluPhe	380
Db	1235	GTTCTGTCAATAAAGTACTGCTAGTTGAACAGCTCTTCCGAAGAGATTCCAAATCATTT	1294
OY	381	LeuLysLysTrpTyrArgValIleGlyLeuSerGlyAspThrGlnLeuLysIleSerPhe	400
Db	1295	TTGAAAGAAATGTTGCTGTTATTGGATTAAGGTGATACCCAACTGAAATATCATTT	1354
OY	401	ProGluValValLysSerCysAspIleIleIleSerThrAlaGlnIleLeuGluAsnSer	420
Db	1355	CCAAAGTTGCAAGTCCCTGTGATTATTATTCAGTACGCTCAATCCTTCATATACACC	1414
OY	421	LeuLeuAsnLeuGluAsnGlyLysValAspAlaGlyValGlnLeuSerAspPheSerLeuIle	440
Db	1415	CTCTTAACTTGGAAATGGAGAAAGATCGTGGTTCATTTCTCAGACTTTTCTTCAATT	1474
OY	441	IleIleAspGluCysHisHisThrAsnLysGluAlaValTyrAsnAsnIleLeuTyrHis	460
Db	1475	ATCATTTGATGAATGATCATCACCAACCAAAAGACGTGTAATATACATCATCTGTCAT	1534
OY	461	TyrLeuMetGlnLysLeuLysAsnAsnArgLeuLysLysGluAsnLysProValIlePro	480
Db	1535	TATTTGATGCGAAGATTGAAAAACATAGCACTCAGAAAAAGAAACCAACTGTTCCGC	1594
OY	481	LeuProGlnIleLeuGlyLeuThrAlaSerProGlyValGlyValAlaIleThrLysIleAla	500
Db	1595	CTTCTCAGATACTGGGACCTAACAGCTTCACTCGGTGTGGAGGGGCGACCAAGTAAACC	1654
OY	501	LysAlaGluGlnHisIleLeuLysLeuCysAlaAsnLeuAspAlaPheThrIleCysThr	520
Db	1655	AAACGTACAGACACACATTTTAAACATATGTGCCAATCTTGATGATTTACTATTAACCT	1714
OY	521	ValLysGluAsnLeuAspGlnLeuLysAsnGlnIleGlnGluProCysLysIleHisAla	540
Db	1715	GTTAAAGAAACCTTGATCAACTAGAAAAACCAATACAGAGCCATGCAGAGATTGGCC	1774
OY	541	IleAlaAspAlaThrArgLysAspProPheLysGlyLysLeuLeuGlnIleMetThrArg	560
Db	1775	ATTGCACTGCAACCAAGAGATTCATTTAAAGAAAACTTCTGAAATATATCAATAGG	1834
OY	561	IleGlnIleThrArgGlnMetSerProMetSerAspPheGlyThrGlnProTyrGlnGln	580
Db	1835	ATTCAAACTATTCTCAAATGAATGCCAATGTAGATTTGGAACTCAACCTATGACAA	1894

QY	1941	ValArgGluAsnLysAlaIleuGlnLysLysCysAlaAspTyrGlnIleAsnGlyIleuLeu	960
Db	2975	GTAAAGCAAAACCAAGCACTGCACAAAGAGCTGCGGCACTATCAATTAATGGTAAATC	3034
QY	961	IleCysLysCysGlyGlnAlaIleTpcIlyThrMetIcvaIHisLysGlyLeuAspLeuPro	980
Db	3035	ATCTCAATATGTGGCCAGGCTTGGGAGCAATATGATGTGCACAAAGGCTTACATTTGGCT	3094
QY	991	CysLeuLysIleArgAsnIleValIlePheLysAsnAsnSerThrLysIysIleTyr	1000
Db	3095	TGTCACAAATTAAGCAATTTTGTACTGCTTTCCAAAAATAATCCAAACAAAGAAATAATAC	3154
QY	1001	LysLysTyrPvalGlnIleuProIleThrPheProAsnLeuAspTyrSerGlyIleLysIleu	1020
Db	3155	AAAAAGCTGGTGGATTAACCTATCATCAATTTCCCAATCTTGACTATTCAGAAATGCTGTTA	3214
QY	1021	PheSerAspGluAsp	1025
Db	3215	TTTACTGATGAGAGT	3229
RESULT 4			
LOCUS	AY075132	3078 bp	linear
DEFINITION	Mus musculus HELICARD mRNA, complete cds.		
ACCESSION	AY075132		
VERSION	AY075132.1	GI:18698980	
KEYWORDS			
SOURCE	Mus musculus.		
ORGANISM	Mus musculus.		
REFERENCE	1 (bases 1 to 3078)		
AUTHORS	Kovacsovics, M., Martignon, F., Michneau, O., Bodmer, J., Hofmann, K., and Tschopp, J.		
TITLE	Overexpression of Helicard, a CARD-Containing Helicase, Cleaved during Apoptosis, Accelerates DNA Degradation		
JOURNAL	Curr. Biol. 12 (10), 838-843 (2002)		
REFERENCE	2 (bases 1 to 3078)		
AUTHORS	Kovacsovics, M., Hofmann, K., and Tschopp, J.		
TITLE	Direct Submission		
JOURNAL	Submitted (24-Jan-2002) Biochemistry, University of Lausanne, ch		
FEATURES	Location/Qualifiers		
Source	1..3078		
CDS	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	1..3078		
	/note="Helicase; contains two CARD domains"		
	/codon_start=1		
	/product="HELICARD"		
	/protein_id="AAL77205.1"		
	/db_xref="GI:18698981"		
	/translation="MSIVCSAEDESFRLNILEFRRLKMYIOVE;VILHLIFLSAEETKE OILKINICNTSAEELLSTLEOGOMPLGWTOMFVELEISGNILARYKPLTIDI PSPSETHADECLILTLTIPTLIDKILINVTGKELGITIEDNRNISAANGSG ESGRELRLRIYCKENNESTLDYLRGTGNALNOELTGSGCPQNTDILASSRDRP AANDCLPAVDESSLETDAMVADIDLEPASTDSSTDSSTISAEQSCFDSISLG NSNNGRSGTGMGSDSDSEVIGTKRSPPELRLPYOMEVAPLIDKNIILCLPTGSS GKTAAVYITKDHLDKKQASGSKVILVANKVMAEOLPFKFNPLYLKKMYRTIGLS GDTQILSEPEVKSVDYIISTADILESLNLSLGGDDGYQIKFSLIIDECNHN KEAYNNIMRRLYKOKLRNNLKNKKMAPLPLPOLIGTASGVGAAGKQSEAKKHIL NICANNIAPFTIKTYKENGKQKHOKIKPKCKPFIADDTRENEKELLETMASTQYIC OKSPSDEGTHTEOMALOMKKAOKDNKRDYCAELHRYNEALQINDIRNIDAMV SHLEFPTDEREKKFAYLNDSDKSDDESSNDCLKGVKSKILHETDEPLMLPDEE NKKMLKLAENPKRYENELIKLRNTLLEOFRSESSSGIIFTKSTAYALQSOMIE NAKRAEVGKRAHLIGAGHSESEVPMTOTEGKEVYSKRTGEIMLILATVAEGLID KECVIIVRGYLTNEIAMVQSGRRARDESTYVLTSSGCVTEKHIIVDEREKMYTK AINVOYNNKPEEVAKHKLLELOVGLITREKKMKVRSIAKOVNDNLSIILCKNCSMLV CSGNHIVTEKMHVNNTPPEKGIYLRKNAKLADADYTNREIILCKGCAAGTAAH VHKQIDLECLKIRNFVNFKNNSPKKOTKKKVELIPRPDILAEVCLYSD"		

ORIGIN

Alignment Scores:

Pred. No.: 1,58e-277 Length: 3078
 Score: 4199.00 Matches: 818
 Percent Similarity: 87.04% Mismatches: 75
 Best Local Similarity: 79.73% Indels: 131
 Query Match: 79.06% Gaps: 2
 DB: 10

US-09-515-363c-2 (1-1025) x AY075132 (1-3078)

QY 1 MetSerAsnGlyThrSerThrAspGluAsnPheArgTyrLeuIleSerCysPheArgAla 20
 DB 1 ATCTGATTTGTTCTGTTCTGACAGACAGCTTTCAGAAATCTATATTCTTACGGCT 60
 QY 21 ArgValLysMetTyrIleGlnValGluProValLeuAspTyrIleThrPheLeuProAla 40
 DB 61 AGGCTGAATATGATACATTCAGTGGAGCCAGCTGCTGACCACTCATCTTCTGCTGTA 120
 QY 41 GluValLysGluGlnIleGlnArgThrValAlaThrSerGlyAsnMetGlnAlaValGln 60
 DB 121 GAAACCAAGAGCAGATTCTTAAAGATCAACACCTGTGTAAACACCAAGTGGAGAA 180
 QY 61 LeuLeuLeuSerThrLeuGlnLysGlyValTTrpIleLeuGlyTTrpThrArgIleVal 80
 DB 181 CTGCTCTGAGCAGCTTGGAGAGGACAATGGCTCTGGATGGAGCAGCATGTTCTG 240
 QY 81 GluAlaLeuArgArgThrGlySerProLeuAlaAlaArgTyrMetAsnProGluLeuThr 100
 DB 241 GAGGGCTTAGAGCAGAGTGGCAATCCCTAGCCGGCGGTATGTCAAACTACACTCAT 300
 QY 101 AspLeuProSerProSerPheGluAsnAlaHisAspGluTyrLeuGlnLeuLeuAsnLeu 120
 DB 301 GATCTCCCTCTCTCTCTCTCTGAGACTGACAGTGTCTCCACTTCTCTGACCTG 360
 QY 121 LeuGlnProThrLeuValAspLysLeuLeuValArgAspValLeuAspLysCysMetGln 140
 DB 361 CTCACCCCACTTGTGTGACAACTTGTGATTAAACAGTGTGTGACATGCTTCTG 420
 QY 141 GluGlnLeuLeuThrIleGlnAspArgAsnArgIleAlaAlaGlnLeuAsnGlnLys 160
 DB 421 AAGGCACTATTGACACTGCAAGACAGAAATCCGATTCTGCTGCGAGAAATACGGGAA 480
 QY 161 GluSerGlyValArgIleLeuLeuLysArgIleValGlnLysGlnAsnTrpPheSerAla 180
 DB 481 GAGTCAGGTGAAGAGAGTGTGACAGAGATTGTCCACAGAGAAACAGTTTCTTAC 540
 QY 181 PheLeuAsnValLeuArgGlnThrGlyAsnAsnGlnLeuValGlnGlnLeuThrGlySer 200
 DB 541 TTCCTGGATGTTCTGCGCAACTGTGAATGATGCACTATTCACAGAACTAACAGCTGA 600
 QY 201 AspCysSerGlySerAsnAlaGlnIleGlnLeuSerGlnValAlaAspCysProGlnVal 220
 DB 601 GGTGCGCCAGAAACACACACAGACTGGCTAACTGCTCCACAGAGATGACCTGAG 660
 QY 221 GluGlnLeuLeuSerThrThrValGlnProAsnLeuGlnLysGlnValTTrpIleMet 240
 DB 661 AATGATGTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
 QY 241 GluAsnAsnSerSerCysLeuSerPheAlaAspSerSerValValSerGlnSerAspThr 260
 DB 721 GGGGCAATATTACCAAGAGCTTCTGTACAGATTCCTGTGTGACCAACAAATCAACACA 780
 QY 261 SerLeuAlaGlnGlySerValSerCysLeuAspGluSerLeuGlnHisAsnSerAsnMet 280
 DB 781 AGTTTGGAGAAAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
 QY 281 GlySerAspSerGlyThrMetGlySerAspSerAspGluLysValAlaAlaArg 299
 DB 841 GCGAGGATTCAGAGCAGCATGGAAAGTATGATGATGAAAGTGTATCCACACAAAGACA 900
 QY 300 AlaSerProGluProGluLeuGlnLeuArgProTyrGlnMetGluValAlaGlnProAla 319

DB 901 GTATCCCGGAGCCAGAACTGAGCTCAGGCTTACCAAAATGAAATGGCTCAATTAAT 960
 QY 320 LeuGlnGlyLysAsnIleIleCysLeuProThrGlySerGlyLysThrValAla 339
 DB 961 CTAGATGGGAAGAAATATATATATATATGCTCCCGACGGGAGATGGGAAACCAATGCT 1020
 QY 340 ValTyrIleAlaLysAspHisLeuAspLysLysLysAlaSerGluProGlyVal 359
 DB 1021 GTTACATCACCAAGAACCTTACAGCAAGAAAGAGGAGCATCTGAATCCGCAAACTT 1080
 QY 360 IleValLeuValAspLysValLeuValGlnGlnLeuPheArgLysGluPheGlnPro 379
 DB 1081 ATCGTTCTTGAATACAGATATGATGACAGAACACTTTCCGAAAGAGTTCACATCA 1140
 QY 380 PheLeuLysLysTTrpTyrArgValIleGlyLeuSerGlyAspThrGlnLeuLysIleSer 399
 DB 1141 TATTGAAAGAAATGATATGATATTTGATTTAATGCGCATACCACTGAAATATATCA 1200
 QY 400 PheProGluValValLysSerCysAspIleIleIleSerThrAlaGlnIleLeuGlnAsn 419
 DB 1201 TTTCCAGAAATGTTCAAACTTACAGATGTTATTCACACACTGCTCAAAATCTTAAAC 1260
 QY 420 SerLeuLeuAsnLeuGlnLysGlnLysGluAspAlaGlyValGlnLeuSerAspPheSerLeu 439
 DB 1261 TCCCTCTTAAATCTGAGAGTGGAGAGATGACGAGTGTGAGCTGTGACACTTCTCTG 1320
 QY 440 IleIleIleAspGluCysHisIleThrAsnLysGlnAlaValTyrAsnAsnIleMetArg 459
 DB 1321 ATTATCATTTGATGATGCTGACATGACACCAAGAGGAGAGTATTAACATCATCAATGA 1380
 QY 460 HisTyrLeuMetGlnLysLeuLysAsnAsnArgLeuLysLysGlnAsnLysTrpValIle 479
 DB 1381 CGATATTGAGGACCAACCTGAGAAACATGACCTCAAGAAACCAAAACCAACCTATT 1440
 QY 480 ProLeuProGlnIleLeuGlnLeuThrAlaSerProGlyValGlnGlyAlaThrGlySer 499
 DB 1441 CCCCTGCGGAGATACTGAGAGTGCAGCTTCACTGTTGTGGAGCAGCCAAAGAG 1500
 QY 500 AlaLysAlaGlnGlnHisIleLeuLysLeuCysAlaAsnLeuAspAlaPheThrIleCys 519
 DB 1501 TCTGAGGCTGAAAAACATATTTAATATATGCTGCAATCTTGAATTCATTAAT 1560
 QY 520 ThrValLysGluAsnLeuAspGlnLeuLysAsnGlnIleGlnGluProCysLysLysPhe 539
 DB 1561 ACAGTGAAGAGAAATCTGTCACACTCAACACCAAAATGAAGAACCTTCGACAGAAAT 1620
 QY 540 AlaIleAlaAspAlaThrArgGluAspProPheLysGlnLysLeuGlnIleMetThr 559
 DB 1621 GTGATTGTGATGACACACAGAAATCCATTTAAAGAAACTTGAAGAAATTAAGCA 1680
 QY 560 ArgIleGlnThrTyrCysGlnMetSerProMetSerAspPheGlyThrGlnProTyrGln 579
 DB 1681 AGCATTCAGACTTACTGCGCAAAAGTCCCATGTCAGATTTTGGACCCAAATTAATAG 1740
 QY 580 GlnTrrAlaIleGlnMetGlnLysLysAlaAlaLysLysGlnLysAsnArgLysGluProVal 599
 DB 1741 CAGTGGCGCATTCAAATGAGAAAAAGCTGCTAAAGCGGAAATTCGCAAGATGAGCT 1800
 QY 600 CysAlaGlnHisLeuArgLysTyrAsnGlnAlaLeuGlnIleAsnAspThrIleArgMet 619
 DB 1801 TGTGAGAACATTTGAGAGACAGACAGAGCCCTACAAATCAACAGACCATCCCATG 1860
 QY 620 IleAspAlaTyrThrHisIleGlnIleThrPheTyrAsnGlnGlnLysAspLysLysPheAla 639
 DB 1861 ATGATGATATAGCCAGCTGAGACATTTACACTGATGAGAAAGAAAGAAACATGTCGA 1920
 QY 640 ValIleGluAspSerSerAspGluGlyLysAspGlnTyrCysAspGlyAspGluAsp 659
 DB 1921 GTCTC---AATGACAGGACAGAGAGTATGACAGGCGACAGCTTGCATTAATGACAAATT 1977
 QY 660 GluAspAspLeuLysLysProLeuLysLeuAspGlnTrrAspArgPheLeuMetThrLeu 679

Db 1978 AAGGGCGATGTAAGAAATCTTGAACCTGAGCAACCAATTCATGATTTG 2037
 QY 680 PhePheGluAsnAsnLysMetLeuLysArgLeuAlaGluAsnProGluLysGluAsn 699
 Db 2038 TTTCTTTGATTAACAGAAATTTGAAAAAAGCTAGCAAAACCAATTCAGATCA 2097
 QY 700 LysLeuThrLysLeuArgAsnThrLysMetGluGlnTyrThrArgThrGluLysLeu 719
 Db 2098 AAACATCAATTAATTAAGAAACAGATCTGAAACATTCACAAAGCTCTCGAGTCTCTC 2157
 QY 720 ArgGlyIleIlePheThrLysThrArgGlnSerAlaTyrAlaLeuSerGlnIleThr 739
 Db 2158 CAGGAAATTAATTTTCAAAAACACACAGACACCTTACACACTTCCAGTGATCATG 2217
 QY 740 GluAsnGluLysPheAlaGluValGlyValLysAlaHisLeuIleGlyValGlyHis 759
 Db 2218 GAAATGCAAAAGTTTGGAGAGTGGAGTCAAAAGCCATCCTGATTCGCGGGGCGAC 2277
 QY 760 SerSerGluPheLysPheMetThrGlnAsnGluGlnLysGluValLysLysPheArg 779
 Db 2278 AGAGAGGAAGTCAAGCCCTGACTGACTGAAACAAAGATCATTAATTTGCG 2337
 QY 780 ThrGlyLysLysLeuLeuLeuIleAlaThrThrValAlaGluGluLysLeuAspLys 799
 Db 2338 ACTGGCGAAATTAATCTGCTTATCGCTACAGCGTGGCAAGGAAGCCGATATCAAA 2397
 QY 800 GluGluAsnIleValIleArgTyrGlyLeuValThrAsnIleIleAlaValGlnAla 819
 Db 2398 GAGTGCATTAATTTGATTTGATGCTGTGACAGACAGATACCAATGCTCCAGAGTC 2457
 QY 820 ArgGlyArgAlaArgAlaAspGluSerThrTyrValLeuValAlaHisLysGlySerGly 839
 Db 2458 CGGGGTGAGCCAGACTGATGAAAGCAGTATCTGCTGCTCCAGCAATGCTCCAGCA 2517
 QY 840 ValIleGluHisGluThrValAsnAspPheArgGluLysMetCysTyrLysAlaIleHis 859
 Db 2518 GTTACGAGAGCGAGATTTGATGATTTCCGAGAGAGATGATCATTAAGCTATTAA 2577
 QY 860 CysValGluAsnMetLysProGluGluTyrAlaHisLysLysLeuGluLeuGlnMetGln 879
 Db 2578 CGTGTCAAAACATGAACACAGAGATGATGATTAAGTTTGGAAATTCAGGTGCA 2637
 QY 880 SerIleMetGluLysLysMetLysThrLysArgAsnIleAlaLysHisLysLysAsn 899
 Db 2638 AGATCTCTGAGAAAGAAATGAAGCAAGCAATTCAGAGCAATTAAGCAAT 2697
 QY 900 ProSerLeuIleThrPheLeuCysLysAsnGlySerValIleAlaCysLysGlyLys 919
 Db 2698 CCAATCTTAATACACTTCTCTCCAAAATTTAGCATGTGCTCTGCTGCGAGAAAC 2757
 QY 920 IleHisValIleGluLysMetHisValAsnMetThrProGluPheLysGluLeuTyr 939
 Db 2758 ATCCATGTCATGAGAAATGATCATGTCATTAATGCAACCAATTCAGAGCTCTAC 2817
 QY 940 IleValArgGluAsnLysAlaLeuGlnLysLysCysAlaAspTyrGlnIleAsnGln 959
 Db 2818 ATTGTAGAGAAACAAAGACACTGCAAAAGAAATTTGCTATTATCAGATCAATGAGAG 2877
 QY 960 IleIleCysLysCysGlyGlnAlaTrrpGlyThrMetMetValHisLysGlyLeuAspLeu 979
 Db 2878 ATTATCTGCAAGTGTGCGCGCTTGGGGAACATGATGTGCAAAAGCTTACAGATTG 2937
 QY 980 ProCysLeuLysIleArgAsnPheValValPheLysAsnAsnSerThrLysLysGln 999
 Db 2938 CATTGCTTAAATTAAGAAATTTGTAAGTCAATTCACAAATTAATCACTCCAGAGAACAG 2997
 QY 1000 TyrLysLysTrrpValGluLeuProIleThrPheProAsnLeuAspTyrThrGlyCys 1019
 Db 2998 TACAGAGAGGTGGGATGCTATCAGATTTCTGCTGATTGACTACTAGAAATCTGCG 3057
 QY 1020 LeuPheSerAspLys 1025
 Db 3058 TTGTATAGTATGATGAAGAT 3075

RESULT 5
 AF374384
 LOCUS
 DEFINITION
 Mus musculus strain C57BL/6J melanoma differentiation associated
 gene 5-like protein mRNA, complete cds.
 ACCESSION
 AF374384
 VERSION
 AF374384.1
 KEYWORDS
 GI:20385627
 ORGANISM
 Mus musculus.
 Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE
 1 (bases 1 to 3771)
 Kang, D.-C. and Fisher, P.B.
 Murine homolog of melanoma differentiation associated gene-5 - a
 DEXH group RNA dependent ATPase
 Unpublished
 JOURNAL
 2 (bases 1 to 3771)
 Kang, D.-C. and Fisher, P.B.
 Direct Submission
 Submitted (26-APR-2001) Urology, Columbia University, P & S. 630
 West 168th, New York, NY 10032, USA
 FEATURES
 source
 location/Qualifiers
 1..3771
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 235..3312
 /note="RNA-dependent ATPase: predicted molecular weight:
 115954.50 Da"
 /codon_start=1
 /product="melanoma differentiation associated gene 5-like
 protein"
 /protein_id="AA021359.1"
 /db_xref="GI:20385628"
 /translation="MSIVCSADESFRNLIFFRRLKMYIQVEVLCILFLISAEKTE
 ILKIKNTGCMTSAELSLSTLEOGOMPLGWTOMFVEALHESGKPLARVYPTLD
 PPSSETAHDCILHLLOPTLVKLLINVDLTGCEKGLIYEDNRRIASAGNSGN
 EGVREILRRIVOKMNEFSLFDVIRGTGNAALPOELIGGCEEDNDLVSRRDGP
 AANECILFAVDESSLETANVDDILPEASTDSSTTESITSLARSGSPDSLSLH
 NSNMGRDSGTMGSDSESVIOTKRSPPELQRYOMEVNOFALDKNNTICLPTDS
 GTRVAVYITDQHLDKKQASGVIVLVKVMALAEQLEFKIENPLKMYRIIGLS
 GDTOLKISPEPEVYSYDVIISTAOILSLNLDSGDDGVQIDPSLIIDECNHN
 KEAVNIVRRITLKKOKLRNLDKKNKPAIPLPOLGLTASGVAAKQSGAEKHLI
 NIGANLDAFTTKYKENIGQIKHOKKECKKFFVADTRENPKHLEIWAISTOTC
 OKSPSDGTGHTEDMAIOMERKAKKDKNRDRCAELKRLKALJINDIRIMDAY
 SHLEAFYDERKRRKRAVLNDSKSDSSSCNDLKGIVKSKLUBETDERLMLFTD
 NKKMLKLAENKRENERLILRLNTLIDQTFRSESSRGIIKTRQSTVERLMSIMH
 NKAFAVEGVKAAHHLIGAGHSESEVKPMTOEKEVYSKRTGEI NLIATVAEGGLD
 KECNIVIRGIVLNETLAWQARARARADESTYLVTSGSQVTEKEIYNDRERKMK
 AINRQNNKPREYAKITILEOVOSTLEKKMKVKSIAQVNDNESLIITLCKNCKMY
 CSGENIHVIEKMHVNMTPERRGLTIVENALOKKRFADYOTN EITCKCGAANGTMH
 VHKGLDPLCLIRINRVNFKNSPKKQYKKVVELPIREPLDNDSEYLYSDSD"
 1228..1245
 /note="ATP binding motif"
 1564..1575
 /note="DEXH helix signature: unclassified site"
 3753..3760
 misc_feature
 misc_feature
 polyA_signal
 BASE COUNT 1183 a 821 c 866 g 901 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 4.4e-277 Length: 3771
 Score: 4194.00 Matches: 817
 Percent Similarity: 86.94% Conservative: 75
 Best Local Similarity: 79.63% Mismatches: 132
 Query Match: 78.97% Indels: 2
 DB: 10 Gaps: 2
 US-09-515-363c-2 (1-1025) x AF374384 (1-3771)
 QY 1 MetSerAsnGlyTyrSerThrAspGluAsnPheArgTyrLeuIleSerCysPheArgAla 20

```
Db 235 ATGTGATTTCTGTTCTGACAGAGAGAGCTTACAGAAATTCATTCTTCAGAGCTC 294
|||||
QY 21 ArgValLysMetThrIleGlnValIleGluProValLeuAspIleThrPheLeuProAla 40
|||||
Db 295 AGGCTGAATATGATACATTCAGTGGAGAGCCAGCTGCTGACACTCATCTTTCGTGTGA 394
|||||
QY 41 GluValLysGlnIleGlnIleThrValAlaThrSerGlyAsnMetGlnAlaValGln 60
|||||
Db 355 GAAACCCAAAGAGCATTTCTTAAAGATCAACCTGTGTATCAACCAATGCGGAGAA 414
|||||
QY 61 LeuLeuLeuSerThrLeuGlnLysGlyValIleThrIleLeuGlnIleThrPheThrAla 80
|||||
Db 415 CTGCTGCTGACACCTTGGAGCAAGGCAATGGCTGTGTGATGAGAGAGATGTTCG 474
|||||
QY 81 GluAlaLeuArgIleThrGlySerProLeuAlaAlaArgIleMetAsnProGluLeuThr 100
|||||
Db 475 GAGGCCCTAGACAGAGTGGCAATCCCTTAGCCGCCGCTATCTCAAACTGACACTGACT 514
|||||
QY 101 AspLeuProSerProSerPheGlnAsnAlaHisAspGluThrLeuGlnIleLeuAsnLeu 120
|||||
Db 535 GATCTGCTGCTGCTCTCTCTCTGAGACTGCCCATGAGAGAGTGTCCACTTGTGACCTTC 594
|||||
QY 121 LeuGlnProThrLeuValAspLysLeuValAlaAspPheLeuAspLysGlyMetGln 140
|||||
Db 595 CTCACAGCCACTTGGGAGCAACCTGTGATTAACGATCTCTTGACAATGCTTCGAG 654
|||||
QY 141 GluGlnLeuLeuThrIleGlnAspArgAsnArgIleAlaAlaAlaGlnAsnAsnGln 160
|||||
Db 655 AAGGACTATTGACAGTGCAGAGCAAGCAAAATCGGATTTCTGCTGACAGAAATACGGGAT 714
|||||
QY 161 GluSerGlyValArgGlnLeuLeuLysArgIleValGlnLysGlnAsnThrPheSerAla 180
|||||
Db 715 GACTGAGCTGTAAGAGAGCTGTGAGAGAGATGTGACAGAGGAAACTGTTCTTCTACG 774
|||||
QY 181 PheLeuAsnValLeuArgGlnThrGlyAsnAsnGlnLeuValGlnIleLeuThrGlySer 200
|||||
Db 775 TTCTGTGATTTCTGCGCCCAACTGGAATGATGCACTATTCACAAATCAACAGTGTGA 814
|||||
QY 201 AspGlySerGlnSerAsnAlaGlnIleGlnAsnLeuSerGlnValAspIleProGlnVal 220
|||||
Db 835 GCGTGGCCAGCAACAGCAACAGCACTTGGCACTGCTGACAGAGATGATGCTGAGCT 894
|||||
QY 221 GluGlnIleLeuLeuSerThrValGlnProAsnLeuGlnLysGlnValIleTrpGlyMet 240
|||||
Db 895 AATGACTGTCTGCTGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATG 954
|||||
QY 241 GluAsnAsnSerSerGlnSerPheAlaAspSerSerValValSerGlnLeuSerAlaThr 260
|||||
Db 955 GACGACATATTACAGAGGCTTTCTTACAGATTTCTGCTGTGATACAGATTCAGACACA 1014
|||||
QY 261 SerLeuAlaGlnLysSerValSerGlyLeuAspGlnSerLeuGlnHisAsnSerAsnMet 280
|||||
Db 1015 AGTTGGCAAGCAAGTGTGACCTGCTGATGAAAGCTTGTGACATATGAAAGCAACATG 1074
|||||
QY 281 GlySerAspSerGlyThrMetGlySerAspSerAspGlnGluAsnValIleAlaAlaArg 299
|||||
Db 1075 GGGAGGAGTTCAGGCACTGAGAGAGTGTGATGATGATGATGATGATGATGATGATGATG 1134
|||||
QY 300 AlaSerProGlnProGlnLeuGlnIleuArgProGlnIleMetGlnValGlnIleProAla 319
|||||
Db 1135 GTATCCCGGAGCCAGAACTGCACTGAGGCTTACCAATGATGATGATGATGATGATGATG 1194
|||||
QY 320 LeuGlnGlyLysAsnIleIleIleCysLeuProThrGlySerGlnLysThrArgValAla 339
|||||
Db 1195 CTAGATGGGAGAGATATTATTATCTGCTCCCGGAGGAGAGAGAGAGAGAGAGAGAGAG 1254
|||||
QY 340 ValThrIleAlaLysAspHisLeuAspLysLysLysAlaSerGlnProGlnLysVal 359
|||||
Db 1255 GTTACATCAACCAAGATCACTTAGACAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAG 1314
|||||
QY 360 IleValLeuValAsnLysValLeuLeuValGlnIleuPheAsnGlnIlePheGlnPro 379
|||||
Db 1315 ATGCTTTCTGTCATTAAGTAAATGTTAGCAGAACACTTTCCGAAAGAGTTTAACGCA 1374
|||||
QY 380 PheLeuLysLysThrPylArgValIleGlyLeuSerGlyAspThrGlnLeuThrIleSer 399
|||||
Db 1375 TATTTCAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1434
|||||
QY 400 PheProGlnValValLysSerCysAspIleIleIleSerThrAlaGlnIleLeuGlnAsn 419
|||||
Db 1435 TTTCGAGAGTGTCAAACTTACGATGTTATTATCAGACAGCTGCTCAAACTCTTAAAG 1494
|||||
QY 420 SerLeuLeuAsnLeuGlnAsnGlnGlnAspAlaGlyValGlnLeuSerAspPheThrSer 439
|||||
Db 1495 TCCCTCTTAATCTGAGAGTGCAGATGATGAGGCTGTGACCTGTGACACTTCTGCTG 1554
|||||
QY 440 IleIleIleAspGlnCysHisHisThrAsnLysGlnAlaValThrAsnAsnIleMetArg 459
|||||
Db 1555 ATTATCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1614
|||||
QY 460 HisThrLeuMetGlnLysLeuLysAsnAsnArgLeuLysLysGlnLysAsnLysProValIle 479
|||||
Db 1615 CGATATTGACAGAGAGTGCAGAAACATGACCTCAAGAAACAAACAAACAAACCAACAT 1674
|||||
QY 480 ProLeuProGlnIleLeuGlnLysLeuThrAlaSerProGlnValGlyValAlaThrLysGln 499
|||||
Db 1675 CCGCTGCGGAGATCTAGACAGTGCAGCTTCCCTGCTGTGAGAGAGAGAGAGAGAGAG 1734
|||||
QY 500 AlaLysAlaGlnGlnHisIleLeuLysLysCysAlaAsnLeuAspAlaPheThrIleLys 519
|||||
Db 1735 TGTGAGCCGAAACATATTATTAATATATGCTGCACTGATGATGATGATGATGATGATGATG 1794
|||||
QY 520 ThrValLysGlnAsnLeuAspGlnLeuLysAsnGlnIleGlnIleProCysLysLysThr 539
|||||
Db 1795 ACAGTGAAAGAGAAATCTGTGCACTCAACCAACCAACCAACCAACCAACCAACCAAT 1854
|||||
QY 540 AlaIleAlaAspAlaThrArgGlnAspPropylLysGlnLysLeuGlnIleuThr 559
|||||
Db 1855 GTGATTTGCTGATGACACCAAGAGAAATCTCTTAAGAGAACTCTGAAATTAAGCA 1914
|||||
QY 560 ArgIleGlnThrTyrcGlnMetSerProMetSerAspPheGlyThrGlnThrGlyGln 579
|||||
Db 1915 ACCATTGACAGCTTACTGCGCAAAAGTCCAAATGTCAGATTTTGGAAACCAATATATGAG 1974
|||||
QY 580 GlnThrPalaIleGlnMetGlnLysLysAlaAlaLysLysGlyAsnArgLysGlnArgVal 599
|||||
Db 1975 CAGTGGCCATTCATTAATGAGAAAAAGCTGCTTAAGAGCGGAAATCGCAAAATATCTG 2034
|||||
QY 600 CysAlaGlnHisLeuArgLysThrArgGlnAlaLeuGlnIleAsnAspThrIleThrMet 619
|||||
Db 2035 TGTGCAAGACATTTGAGAGAGTACACAGAGCCCTACAAATCAACGACAGATTCATGATG 2094
|||||
QY 620 IleAspAlaArgThrHisLeuGlnIleThrPheThrArgGlnGlnLysAspLysHisAla 639
|||||
Db 2095 ATTGATGATATATGACCTGAGGCACTTCTACACTGATGAGAAAGAAAGAAAGAAATGCA 2154
|||||
QY 640 ValIleGlnAspAspSerAspGlnGlyLysAspAspGlnThrCysAspLysAspLysAsp 659
|||||
Db 2155 GTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2211
|||||
QY 660 GluAspAspLeuLysLysProLeuLysLeuAspGlnIleThrAspArgPheLeuThrLeu 679
|||||
Db 2212 AAGGCGGATGTAAAGAAATCTTTGAACCTGAGCGAAAGAGATGATGATGATGATGATGATG 2271
|||||
QY 680 PhePheGlnAsnAsnLysMetLeuLysArgLeuAlaGlnAsnProGlnIleTyrcGlnAsnGln 699
|||||
Db 2272 TTCTTTGATTAACAAAGAAATGTTGAAAAAAGTACGTGAAAAAAGCAAAATATGACAGATGA 2331
|||||
QY 700 LysLeuThrLysLeuArgAsnThrIleMetGlnIleThrArgThrGlnGlnIleSerAla 719
|||||
Db 2332 AAACATCAATTAATTAAGAAACAGTACTGCAACATTCACAAAGTGTGAGAGAGCTGCT 2391
|||||
QY 720 ArgGlyIleIlePheThrLysThrArgGlnSerAlaArgAlaLeuSerGlnThrPheThr 739
|||||
Db 2392 CGAGGAATTAATTTTACAAAAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2451
|||||
```

```

REFERENCE
AUTHORS
1 Ishibashi,T., Kanehori,K., Yosida,M., Watanabe,S., Ishida,S., Ono,Y., Hotta,T., Hirooka,S., Murakawa,K., Takiguchi,S., Kusano,J., Watanabe,M., Fujimori,K., Tanai,H., Ishida,H., Yamashita,H., Chiba,Y., Sugiyama,T., Ito,H., Otsuki,T., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kakiuchi,Y., Saito,K., Nishikawa,T., Kimura,K., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Magetsuna,M., Takahashi-Fuji,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sudo,O.S., Nagahara,K., Masuno,Y., Nagai,T. and Sugiyama,T.
TITLE
NEDO human cDNA sequencing project
JOURNAL
unpublished
REFERENCE
2 (bases 1 to 1776)
AUTHORS
Ishigai,T., Otsuki,T. and Sugiyama,T.
TITLE
Direct Submission
JOURNAL
Submitted (24-OCT-2001) Takao Isegai, Helix Research Institute, Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 24, 612, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT
NEO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan. cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- and 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.
FEATURES
SOURCE
Location/Qualifiers
1. 1776
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NT2R1200685"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/clone_id="NT2R12"
/note="Cloning vector: pME18SP3-mRNA from NT2 neuronal precursor cells treated 2-weeks mitotic inhibitor after 250-weeks retinoic acid (RA) induction. majorly NT2 neuron"
156
/feature="unnamed protein product"
/codon_start=1
/protein_id="BAB7114.1"
/db_xref="GI:16551654"
/translation="MTRIGTCOMSPMSDFSTOPYEOMALMEKKAKEGNREKRYCC EHLKRYNEALQINDTIRMIDAYTHLETFYNEKDKFAVILPDSDEGJDEYCDGDEDEDDKRPKLPKIDPELFTLTFEENKMKLRLEANEPEYENKLTILRTIMQYTRTEESARGIETFTROSAAVALSOMITENKRFVGVCAHHLGLNDSSEKFRPTGQNEQVLSKFRGRINLIATTVAREGDIKRCNVIYRGVLTNINAVYACAGRAKRADESTVLLAHSGGVETRETVNDFERKKMKIKHCVONKPEYAIKILIDWQSTMEKMKTKRIIAKHKNPSSITFLCKMSVLMSCGDIHVEKHHVNMIPFELYIVRNKTLQD KCAVDQIIEITCKGQAMGTVMVHKIGDLPLCKIRNFVVFVKNSIKROYKKWELWELITFLPDVSECCLEFDEED"
BASE COUNT
666 a 291 c 363 g 456 t
ORIGIN
Alignment Scores:
Pred. No.: 2,396-187 Length: 1776
Score: 2874.00 Matches: 546
Percent Similarity: 99.64% Conservative: 3
Best Local Similarity: 99.09% Mismatches: 2
Query Match: 54.11% Indels: 0
Gaps: 0
US-09-515-363C-2 (1-1025) x AK056293 (1-1776)
QY 475 AsnLysPProValIleProLeuProGlnIleLeuGlyLeuThrAlaSerProCysValIleY 494
Db 1 CACCAACCGAGTATTCCTCCCTCAGATACGAGGGGAGCTAACAGGCTTCACCTGATTTGGA 60
QY 495 GAlaAThLrLysGAlaAlaLysAlaGluGluIuHHisIleLeuLysLLeuGlyLysAlaAsn 514
Db 61 GGGGGCCAGCAGCAAGCCCAAGCTGATAGACACATTTTAAACTATGTGCCAATTTTGAT 120

```

QY 515 AlapheThlileuThyValylscLuanleuAspGlnleuLysAsnGlnilectGln 534
 |||||
 Db 121 GAATTACTATTAAACCTTTAAAGAAACCTTGATCAACCTAAACCTAAATACAGAG 180
 QY 535 ProCysLysLysPheAlaIleAlaAspAlaThrArgGluAspProPheGlnGlnLys 554
 |||||
 Db 181 CCATGCAAGAAATTTGCCATTGACATGCAACCAAGAGATCTATTTCAGAGAACTT 240
 QY 555 LeuGluIleMetThrArgIleGlnThrTyrcysGlnMetSerProMetLysAspPheGly 574
 |||||
 Db 241 CTAGAAATATATGCAAGATCAAACTTATGTCTCAAAATGATGTCATGATTTTGA 300
 QY 575 ThrGlnProTyrcysGlnThrAlaIleGlnMetGlnLysValAlaLysGlnLys 594
 |||||
 Db 301 ACCCAACCTATGACAAATGGCCATTCATTAATGAAAAAAGCTGCAAAACAGAGAA 360
 QY 595 ArgLysGluArgValCysAlaGlnIleLysLeuArgLysTyrcysGlnAlaIleGlnLys 614
 |||||
 Db 361 CCGAAGAAAGCGTTTGTTCAGACATTTGAGGAAGTACAAATGAGCCCTACAAATTA 420
 QY 615 AspThrIleArgMetIleAspAlaThrThrlsLeuGlnThrPheTyrcysGlnGlnLys 634
 |||||
 Db 421 GACACAAATTCGAATGATGATGCTATCTCAATCTTGAACCTTCTATATGAGAGAA 480
 QY 635 AspLysLysPheAlaValIleGlnAspAspSerAspGlnGlyGlyAspAspGlnTyrcys 654
 |||||
 Db 481 GATAGAAAGCTTGCATGATAGAGATGATGATGATGATGATGATGATGATGATGAT 540
 QY 655 AspGlyAspGluAspGlnAspLysLeuLysProLeuLysLeuAspGlnThrAspArg 674
 |||||
 Db 541 GATGGTGATGAGATGAGATGATTTAAAGAAACCTTGAACCTGATTAACAGATTA 600
 QY 675 PheLeuMetThrLeuPhePheGlnLysAsnLysMetLeuLysArgLeuValGlnAsnPro 694
 |||||
 Db 601 TTTCATGACTTATTTTAAAAACAATAAATGTTGAAAAAGCTGCGTGAAGAAACCT 660
 QY 695 GlnTyrcysGlnGlnLysLeuThrLysLeuArgAsnThrIleMetGlnGlnThrThrArg 714
 |||||
 Db 661 GATATGAAATGAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 720
 QY 715 ThrGlnLysSerAlaArgLysIleIlePheThrLysThrArgLysSerAlaThrAlaGln 734
 |||||
 Db 721 ACTGAGGATGACGACGAGAGATTAATCTTACAAAACGACGACAGTGTATATGCTCT 780
 QY 735 SerGlnThrIleThrGlnAsnGlnLysPheAlaGlnValGlyValLysAlaHisHisLeu 754
 |||||
 Db 781 TCCCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
 QY 755 IleGlyAlaGlyHisSerSerGlnPheLysProMetThrGlnAsnGlnGlnGlnVal 774
 |||||
 Db 841 ATTGGAGCTGACACGACGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
 QY 775 IleSerLysPheArgThrGlnLysIleAsnLeuLeuAlaThrThrValAlaGlnGln 794
 |||||
 Db 901 ATTAGTAATTTTGGCAGCTGAGAAATTAATCTGCTATGCTGACCACTGGCAGAGAA 960
 QY 795 GlyLeuAspIleLysGlnCysAsnIleValIleArgTyrcysGlnValThrAsnGlnLeu 814
 |||||
 Db 961 GGTCTGGATATTAAGAAATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
 QY 815 AlaMetValGlnAlaArgGlnArgAlaArgAlaAspGlnSerThrThrValIleValAla 834
 |||||
 Db 1021 GCCATGCTGACGCGCGTGTGAGCGCAAGCTGATGAGAGACCTACCTGCTGTGCT 1080
 QY 835 HisSerGlySerGlyValIleGlnLysGlnThrValAsnAspPheArgGlnLysMetLys 854
 |||||
 Db 1081 CACAGTGCTGAGAGATTTGACAGCTGAGAGATTAATGATTTCCGAGAGAGAGATGAT 1140
 QY 855 TyrLysAlaIleHisCysValGlnAsnMetLysProGlnGlnTyrcysAlaHisLysIleLeu 874
 |||||
 Db 1141 TATTAAGATATACATGCTGCTCAAAATATGAAACAGAGAGATGATGCTATTAAGATTT 1200
 QY 875 GlnLeuGlnMetGlnSerIleMetGlnLysLysMetLysThrLysArgAsnIleAlaLys 894

Db 1201 GAATTACGATGCAAAAGTATATGAGAAAGAAATGAAACCAAGACAGAAATATATATGAG 1260
 QY 895 HisTyrcysAsnAsnProSerLeuIleThrPheLeuCysLysAsnCysSerValAlaAla 914
 |||||
 Db 1261 CATTCAGAAATTAACCCATCTCATTAATACCTTCCCTTCCAAAAACCTGACAGCTGT 1320
 QY 915 CysSerGlyGluAspIleHisValIleGlnLysMetHisHisValAsnMetThrProGln 934
 |||||
 Db 1321 TGTCTGGGGAAGATATCATGTAATGAGAAATGATGATGATGATGATGATGATGATGAT 1380
 QY 935 PheLysGlnLeuTyrcysIleValArgGlnAsnLysAlaLeuGlnLysCysAlaAspTyrcys 954
 |||||
 Db 1381 TTCAGAGAACTTACATTTGTAAGAGAAACAAACACCTGCAAAAGAGTGTGCTACTAT 1440
 QY 955 GlnIleAsnGlyGlnIleIleCysLysCysGlyGlnAlaIleArgGlyThrMetMetValHis 974
 |||||
 Db 1441 CAATTAATATGTTGTAATATCTGCAAAATGTGCGCCAGCTTGGGAGACATATGCTGAC 1500
 QY 975 LysGlyLeuAspLeuProCysLeuLysIleArgAsnPheValValPheLysAsnAsn 994
 |||||
 Db 1501 AAAGCTTACATTTGCTGCTGCTCAAAATGAGAAATTTTGTAGTGTGTTTCAAAATTAAT 1560
 QY 995 SerThrLysGlnGlnTyrcysLysThrValGlnLeuProIleThrPheProAsnGlnAsp 1014
 |||||
 Db 1561 TCACAAAGAAACCAATACAAAGAGGCTAGCAATTAATCTATACATTTCCCAATTTGAC 1620
 QY 1015 TyrSerGlyCysCysLeuPheSerAspGluAsp 1025
 |||||
 Db 1621 TATTCAGAAATGCTGTTTATTTAGTATGATGAGAT 1653

RESULT 7
 LOCUS BC025508 2468 bp mRNA linear 5' end 07-AUG-2002
 DEFINITION Mus musculus, clone MGC:38064 IMAGE:5252257, mRNA, complete cds.
 ACCESSION BC025508
 VERSION BC025508.1 GI:19343862
 KEYWORDS MGC.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Scurionathu; Muridae; Mus.
 1 (bases 1 to 2468)

REFERENCE
 1 Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (06-MAR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 CONTACT NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 COMMENT
 Remark: MGC help desk
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Lohrer Hennighausen Ph.D., Robin Humphreys cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (cDN)

DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc.mgc@nih.gov
 Ahter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Boufield, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Lebaspi, R., Maduro, Q.L., Masfello, C., Masker, B., Maslian, S.D., McCloskey, J.C., McQuell, J., Pearson, R., Stantirpop, S., Thomas, P.J., Touchman, J.W., Tsurgeon, C., Vost, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H., and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILN at: <http://img.mbl.edu>
 Series: IRAC plate: 56 Row: c Column: 12
 This clone was selected for full length sequencing because it

LOCUS	AX300841	1443 bp	DNA	Linear	PAT 30-NOV-2001
DEFINITION	Sequence 10 from Patent WO0185955.				
ACCESSION	AX300841				
VERSION	AX300841.1	GI:17382119			
KEYWORDS					
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eumariota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
AUTHORS	1				
TITLE	Bahr, G., Cocude, C. and Capron, A.				
JOURNAL	Rh16 polypeptides and its fragments and polynucleotides encoding said polypeptides and therapeutic uses				
FEATURES	Patent: WO 0185955-A 10 15-NOV-2001; 1stac (FR) ; INSTITUT PASTEUR DE LILLE (FR)				
source	Location/Qualifiers				
	1..1443				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
BASE COUNT	508 a 261 c 313 g 360 t				1 oth-rs
ORIGIN					
Alignment Scores:					
Pred. No.:	5,84e-158	Length:	1443		
Score:	2442.00	Matches:	475		
Percent Similarity:	99.38%	Conservative:	2		
Best Local Similarity:	98.96%	Mismatches:	3		
Query Match:	45.98%	Indels:	0		
DB:	6	Gaps:	0		
US-09-515-363C-2 (1-1025) x AX300841 (1-1443)					
QY 174	LysGIuAsnTrpPheSerAlaPheLeuAsnValLeuArcJIntHrGIyA:uAsnJleu				193
Db 2	AAAGAAACCTGCTTCCTCGCATTTCTGAATGTTCTTCGTAAACAGGAACAAATCACTT				61
QY 194	ValGIuLeuLeuThrGIySerAspCysSerGIuSerAsnValaGIuLeuLeuSer				213
Db 62	GTCCAAGATTAAACAGGCTCATGTCACAGAAACAACTACAGACATTGACCAATTATCA				121
QY 214	GIuValAspGIyProGIuValaGIuGIuLeuLeuSerThrThValaGIuProAsnLeu				233
Db 122	CAAGTTGATGGCCCAAGTGGAAAGGAACTTTCTTCAACCCAGTTCACCCAAATCTG				181
QY 234	GIuLysGIuValaTrpGIyMetGIuAsnAsnSerSerGIuSerSerPheAlaAspSer				253
Db 182	GAGAAAGAGGCTTGCGGATGGAGAAATACATCATACAAATCTTTGCAATCTTCT				241
QY 254	ValValSerGIuSerAspTrpSerLeuAlaGIuGIySerValSerCysLeuAspGIuSer				273
Db 242	GTAGTTTCAAGATCAACACAAAGTTTGGCAGAAAGAAAGTGTCACTGCTGTACATGAAGT				301
QY 274	LeuGIyHisAsnSerAsnMetCysSerAspSerGIyThrMetGIySerAspSerAspGIu				293

Db	302	CTTGACACTAACACGACACATGGGACAGTGGATTCACGCCACTGGGAAGTGGATTCAGTATTA	361
QY	294	GIuAsnValAlaAlaArgAlaSerProGluLeuGluLeuLeuArgProTyrGluMet	313
Db	362	GAGAAITGGCGACGACAGAGCATCCCGGACGCCAGAACTCCACGCTTACGCGCTTACCAATG	421
QY	314	GIuValAlaGluProAlaLeuGluGlyLysAsnMetLeuLeuSerLeuProThrGluSer	333
Db	422	GAAGTTGGCCGACGCGCTGGAGGGAAGAAATATCATCTTCGCTCCCTCATCAAAAGT	481
QY	334	GIuYsTrpArgValAlaValTyrLeuAlaLysAspHisLeuAspLysLysLysLysAla	353
Db	482	GGAAGAAACCCAGAGTGGCTGTTTACATTTGCCAGATCATTACACAGAGAAAGAAAGCA	541
QY	354	SerGluProGlyLysValLeuValLeuValAsnLysValLeuLeuValGluGluLeuPhe	373
Db	542	TCTGAGCCTGGAAAAAGTATAGTTCTGTCAATAGGTACTGCTAGTTGACAGCTCTC	601
QY	374	ArgLysGluPheGluProPheLeuLysLysTyrTyrArgValLeuLeuLeuSerTrpLysP	393
Db	602	CCCAAGAGGATCCCAACCATTTTGGAAAGAAATGATCTGTATTGTGGATTAAATCTGAT	661
QY	394	ThrGluLeuLysHisSerPheProGluValValLysSerCysAspLeuLeuLeuSerTr	413
Db	662	ACCCACATGAAAAATATCTTCTCCAGAAAGTTGTCAAGTCCGTGTATATTATATATACA	721
QY	414	AlaGluLeuGluAsnSerLeuLeuAsnLeuGluAsnGlyLysAspAlaValGlu	433
Db	722	GCTCAAACTCTGAAAACTCCCTCTTAACTTGGAAATAGCAAGATGCTGCTTCA	781
QY	434	LeuSerAspPheSerLeuLeuLeuAspGluCysHisHisThrAsnLysGluAlaVal	453
Db	782	TTGTAGACTTTTCTTCATTATATATGATGAAGATCATACCCAAACAAACAGTAGT	841
QY	454	TyrAsnAsnHisMetArgHisTyrLeuMetGluLysLeuLysAsnAsnArgLysLys	473
Db	842	TATATATACATCATGAGGCATATTATTATGTCAGACAGTTGAAAAACATACACTCAGAAA	901
QY	474	GluAsnLysProValLeuProLeuProGluLeuGluLysLeuThrAlaSerProGlyVal	493
Db	902	GAAGAACAAACCAAGATATCCCTCCCTCCTTCATCATCTGAGACTAACACTTCATCTGT	961
QY	494	GlyGlyAlaThrLysGluAlaLysAlaGluGluHisHisLeuLysLeuCysAlaAsnLeu	513
Db	962	GGAGGGCGCCAGAGCAAGCCCAAGCTGGAAGAACATTTTAAACTATGTCACATCTT	102
QY	514	AspAlaPheThrHisLysThrValLysGluAsnLeuAspGluLeuLysAsnGluLeuGln	533
Db	1022	GATGCATTTACTATTAAACTGTTAAAGAAAACCTTGATTAACCTGAAAAACATAATACAG	108
QY	534	GluProCysLysLysPheAlaHisLeuAlaAspAlaThrArgLysAspProPheLysLys	553
Db	1082	GAGCATCTCAGAGAGTTGGCATTTGGCAGATGCAACGACAGACATGCATTAAACGAAA	1144
QY	554	LeuLeuGluLeuMetThrArgLeuGluThrTyrCysGluMetSerProMetSerAspPhe	573
Db	1142	CTTCTAGAGAAATATGACAGAGATTCAAACTATTATTGTCAAATGAGTCCAAATGTCATTTT	120
QY	574	GlyThrGluProTyrGluGluThrAlaHisGluMetGluLysLysAlaHisLysGly	593
Db	1202	GCAACTCAACCTATGAACAATGGGCACTTCAATGGAAAAAAACCTCCAAAGACAGCA	126
QY	594	AsnArgLysGluLysValCysAlaGluHisHisLeuArgLysTyrAsnGluAlaLeuHisLeu	613
Db	1262	AATGCCAAAGAAAGTGTGTTGGCAAGAACATTTGAGGAAGACATTAAGGCTTAAAT	132
QY	614	AsnAspThrIleArgMetIleAspAlaTyrThrHisLeuGluThrPheTyrAsnGlu	633
Db	1322	AATGACACAATTGCAATGATAGATGGCTATACCATCTTCAAACTTTCTATATATAGAG	138
QY	634	LysAspLysLysPheAlaValLeuLysAspAspSerAspGluGlyLysAspLysLys	653

DB 1382 AAGATAGAGTTCGACATAGAGATGATGAGGTGGTGGTATGATGATAT 1441

RESULT 9

LOCUS BC004031 2050 bp mRNA linear ROD 07-AUG-2002

DEFINITION Mus musculus, clone IMAGE:3495361, mRNA, partial cds.

ACCESSION BC004031

VERSION BC004031.1 GI:13278455

KEYWORDS

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 2050)

AUTHORS Strausberg, R.

TITLE Direct Submission

JOURNAL Submitted (28-FEB-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT

Contact: MGC help desk

Email: cgapps-r@mail.nih.gov

Tissue Procurement: Gilbert Smith, Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)

DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305

Web site: <http://www.shgc.stanford.edu>

Contact: (Dickson, Mark) mdickpaxil@stanford.edu, Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: <http://image.lnll.gov>

Series: IRAC plate: 8 Row: K Column: 18

This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

FEATURES

source

location/Qualifiers

1..2050

/organism="Mus musculus"

/db_xref="taxon:10090"

/map="C28281"

/clone="IMAGE:3495361"

/issue_type="Mammary tumor metastasized to lung. Tumor arose spontaneously from a spontaneous normal mammary (clonal) outgrowth infected with the virus MTV."

/clone_lib="NCI CGAP_Lu29"

/lab_host="DH10B"

/note="vector: pCMV-Sport6"

<1..1545

/codon_start=1

/product="Unknown (protein for IMAGE:3495361)"

/protein_id="AA04031.1"

/db_xref="GI:13278455"

/translation="ANDAFITIKYKENGIGLKHQIKG (KKFVIAAD)REMPREKL LEMASIOYTCQSPMSDFGTHTEWMA (OMERKAKK) NKRDV AELUKYNELO INDITIMIDAYSHLEFYDEKREKPAVINDSDDEPSSNDOLGKVKSLKIDE TDEFLNLFENKKMLKLAENPKYENKILKLRNIILOFTRESSESSGIIITITRO STVALSOMIMENAKFAEVGVAKHILGAGHSSSEKPMTLEPKFVSKRTEINILITTVTAEEGLDKEONCVIRYGLVINEIAVQARARA IETVYVTSNGSVTEERI VVDFREKMYKAIINRYQNNKPREYAKLLELVQSLPDKKKYKSTIAQYNDNEBELI TLCKKCSMLVSGENIHVIEKMEHNHNPETRGVITVINKALOKKEDVDTNGEII CCGGAMGTMMVHKGLDPLCKIRNFVNFKNNSPKKYKKKVELPIRPDDYSEYCLYSDSD"

CDS

BASE COUNT 747 a 374 c 427 g 502 t

ORIGIN

Alignment Scores:

Pred. No.: 3,87e-146 Length: 2050

Score: 2271.50 Matches: 438

Percent Similarity: 90.87% Conservative: 30

Best Local Similarity: 85.05%

Query Match: 42.77%

DB: 10 Gaps: 1

US-09-515-363C-2 (1-1025) x BC004031 (1-2050)

511 ALaanleuAspAlaPheThrIleLysThrValLysGluAsnLeuAspGlnLeuLysAsn 530

|||||

1 GCCAATCTTGATGCTTACCATTTAAACAGTGAAGAAATCTTGCTCACTTAAACAC 60

531 GlnIleGlnIleProCysLysLysPheAlaIleAlaAspAlaThrArgGluAspProIle 550

|||||

61 CAATTAAGCAACCATGCAAGAAATTTGATGCTGATGATACAGAGAAATCAAT 120

551 LysGluLysLeuLeuGlnIleMetThrArgIleGlnIleThrCysGlnMetSerProMet 570

|||||

121 AAAGAGAAACCTAGTAAATTTGGCAGCATTTGACCTTACGCAAAAAGTCAATG 180

571 SerAspPheGlyThrGlnProTyrGlnIleGlnIleThrAlaIleGlnMetGluLysAlaAla 590

|||||

181 TCAGATTTGGAAACCCACCATTTATGACGACGTGGCCATTTCAATGAGAAATAA 240

591 LysLysGlyAsnArgLysGlnArgValCysAlaGlnLysLeuArgLysTyrAsnProIle 610

|||||

241 AAAGAGGAAATCGCAAGATGCGCTGTGGCAAAATTTGAGGAAATGACCAACCC 300

611 LeuGlnIleAsnAspThrIleArgMetIleAspAlaThrIleLysGlnIleThrIleArg 630

|||||

301 CTACAAATTCACGACGCGATCGCAATGATGATGCTGATAGCCACTGGACACAATAC 360

631 AsnGlnLysLysAspLysLysPheAlaValIleGlnAspAspSerAspGluLysLysAsp 650

|||||

361 ACTGATGAGAAAGAAAGAAAGATTCGACGCTCC---AACGACGACGACGACGACG 417

651 AspGluTyrCysAspGlyAspGlnLysGlnLysAspLysLysProLeuLysLeuAsp 670

|||||

418 GAGCCGACGAGTGGCAATGACCACTTAAGGGCGATGTAAACAAATCTTTGAAAT 477

671 GlnThrAspArgPheLeuMetThrLeuPhePheGlnAsnAsnLysMetLeuLysAlaLeu 690

|||||

478 GAACGAGATGAATTTCTCTGTAATTTGTTCTTGTATACAAAGAAATGTTGAATAACATA 537

691 AlaGluAspProGlnIleTyrGlnLysGlnLysLysLeuThrLysLeuArgAsnThrIleMetGlu 710

|||||

538 GCTGAAAACCCAAATACGAGATGAAATGAAATCACTTAATTAAGAACGAGATATGAGAA 597

711 GlnTyrThrArgThrGlnLysLeuSerAlaArgGlyIleIlePheThrLysThrArgIleSer 730

|||||

598 CAATTCACAAAGCTGACGAGCTCCGACGAAATATTTTCACAAAACGACGACGAGAC 657

731 AlaTyrAlaLeuSerGlnIleThrIleThrGlnLysGlnLysPheAlaGlnValGlyValLys 750

|||||

658 ACCTACGACCTTCCGACGATCATGAAATGCAAAATTTGGGAAATTTGATGATGATG 717

751 AlaHisIleLeuIleGlyAlaGlnIleSerSerGlnPheLysProMetThrGlnAsnIle 770

|||||

718 GCGCATCACTGATTTGGCGCGGCGACACAGATGAAGTCAACCCATGCTACAGATGATA 777

778 CAATAAGAGTCACTTGTAAATTTCCACTGGCGGAATTAATCTGTTATCTGATACGACG 837

791 ValAlaGlnGlnGlnLysLeuAspIleLysGlnLysAsnIleValIleArgTyrGlyLeuVal 810

|||||

838 GTGGCAGAGAGAAAGCTGTGATATCAAGAGTGAATATTTGTTATCTGATATGACCTTATC 897

811 ThrAsnGlnIleAlaMetValGlnAlaArgGlyArgAlaArgAlaAspLysSerThrIleArg 830

|||||

898 AGCAATGAGATGACCATGCTGCTCCAGCGCGGCGTGGAGCCAGACGATGATAAAGACAGTAC 957

831 ValLeuValAlaHisSerGlySerGlyValIleGlnHisGlnThrValAsnAspPheArg 850

|||||

958 GTCCGTGTACCGACGAGTGGCTGAGAGATTACGAAAGGAGGAGATGTTAATGATTTCCGAA 1017

QY	292	ASPLGIUASNVALAIAIALARALBSESRPROGILUPROGILUEGLNLUARPRCTF	311
Dd	671	GATGAAGAAATGTGGCGACCAAGACATCCCCGGGCGCAGAATCCAGCTTAC	730
QY	312	GLINELGUVALAIAGINPROALAEUGLUDLYLSASNIIELEIEECYLEUPROTFR	331
Dd	731	CAATGGAAGTTGCCCGACGCCAGCCTTGSAAGSAAATATCATCATCTGTCCCTACA	790
QY	332	GLYSERGLYSTRIRATQVALLAVALTyrILLEAlalysASPHisLEUAFLYSLYLS	351
Dd	791	GGGAGTGGAAAAACCAGAGTGGCTGTTTACATTGGCCAAGATTCACTTAGAACAGAAA	850
QY	352	LYSILASERGUPROGILYLYSVAILLEVALLLEVALLSNLYSVALLLEULVEIAGLUG	371
Dd	851	AAGCATCTGAGCCCTGGAAAAAGTTATACTCTTGTCATAAGTAACCTCTAGTTGAACAG	910
QY	372	LEUPHEARTLYSGIUPHGNIPROPhELEULYSYSTPYRATQVALIIIGLYLEUSER	391
Dd	911	CTCTTCGCAGAGGAGTTGCCACACATTTTTGAAGAAATGATTCGTATTGGATTAGT	970
QY	392	GLYASPHRTGLILEULYSIIIESERPHEPROGLUVALVALYSSERCYASPIELLEILE	411
Dd	971	GGTGATACCCCAACGAAAAATATCATTTCCACAAGTTCGAAGTCTCTGTGAATTATATTC	1030
QY	412	SERPIRIAGIINIIIEUGLIJASNSERLEUNSNLEUGIWAANGIUGIIVASPLACTLY	431
Dd	1031	AGTACAGCTCAAATCTTGAAAACTCCCTCTTAAACTTGCAAAATGGAGAGATCTCTCT	1090
QY	432	VALGINLEUSERASPHERSERLEUIELLEIEASPCILUCYSIHSHISTIASNLVSGLU	451
Dd	1091	GITCATTTGTACAGACTTTCCCTCATTTATCATGTGATGATGATGATGATGATGATGATG	1150
QY	452	ALAVALTYRASNANILLMELETARGHISTYRLEUMETGLINLYSLEULYSALASINARGLU	471
Dd	1151	GCAGGTATATATACATCATGAGCATATTGATGACGAAGTTGAAAAAANAANAANAANA	1210
QY	472	Lys 472	
Dd	1211	AAA 1213	
RESULT 11			
LOCUS	AX300838	1284 bp	DNA linear PAT 30-NOV-2001
DEFINITION	Sequence 7 from Patent WO0185955.		
ACCESSION	AX300838		
VERSION	AX300838.1	GI:17382116	
KEYWORDS	human.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eumariota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	Bahr,G., Cocude,C. and Capron,A.		
JOURNAL	Rhl16 polypeptides and its fragments and polynucleotides encoding		
FEATURES	said polypeptides and therapeutic uses		
source	Patent: WO 0185955-A 7 15-NOV-2001;		
	Istac (FR); INSTITUT PASTEUR DE L'ILE (FR)		
	Location/Qualifiers		
	1..1284		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
BASE COUNT	489 a 198 c 261 g 335 t	1 oth r s	
ORIGIN			
Alignment Scores:			
Pred. No.:	1.68e-127	Length:	1284
Score:	1995.00	Matches:	378
Percent Similarity:	99.47%	Conservative:	0
Best Local Similarity:	99.47%	Mismatches:	2
Query Match:	37.56%	Indels:	0
DB:	6	Gaps:	0

QY 1006 LeuprolidehphroksanleuaspTYrSerGluGlycysleuPhesrTAspGluasp 1025
 DB 1082 TTACGATACATTTCCCATTTGACTATTCAGATGCTGTTATTATTATGATGAGAT 1141
 RESULT 12
 AF316999
 LOCUS AF316999 2326 bp mRNA linear R00 04-MAR-2001
 DEFINITION Mus musculus D1119p2 mRNA, complete cds.
 VERSION AF316999
 KEYWORDS AF316999.1 GI:13194585
 SOURCE Mus musculus.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 2326)
 AUTHORS Miyoshi, K., Cui, Y., Riedinger, G., Robinson, P., Jachocky, J.,
 Zou, L., Oka, T., Dewar, K. and Hennighausen, L.
 TITLE Structure of the mouse Stat 3/5 locus: evolution from Drosophila to
 zebrafish to mouse
 JOURNAL Genomics 71 (2), 150-155 (2001)
 MEDLINE 21100887
 PUBMED 1161808
 REFERENCE 2 (bases 1 to 2326)
 AUTHORS Cui, Y., Zhang, M., Hanover, J. A. and Hennighausen, L.
 TITLE Identification of the Stat3/5 locus that encode novel
 proteins located in the endoplasmic reticulum and the cytoplasm
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 2326)
 AUTHORS Cui, Y. and Hennighausen, L.
 TITLE Direct Submission
 JOURNAL Submitted (24-OCT-2000) Laboratory of Genetics and Physiology,
 National Institutes of Health, 8 Center Drive, Bethesda, MD 20892,
 USA
 FEATURES
 source
 1. 2326
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="11"
 /map="next to Gcn5"
 1. 2326
 /gene="D1119p2"
 2. 2038
 /gene="D1119p2"
 /product="unknown"
 /protein_id="AAK15474.1"
 /db_xref="GI:13194586"
 /translation="MELRPQOMELTPALDEKNIIIMTGAGCTKAAPVAKKHLET
 VDRGVVYLVNRYVHLSOHAEEFRMLDKHPTVITLSGKSRGFGIMARSHDLIC
 TAEILQIALNSSEDEHELEEFSLIVDECHTKKDTVNTLSRYLEOKLRKEPL
 PVDLGLTASPGTGATKLGALIDHILQICANIPTGINSFNCYSOLIMHPKPKQY
 DLQORRADPFGDLIKKILNQIHQOLEMPDLKQFGTMYEQVAVQLEKDAEGLQE
 ORVYHLIRYNDALFIDHTVARADALMDIYDREKTKTQVRAESMLLKIFDHC
 KNYGOLARGPENPKLEMLERILLKQCSHCHLFTKTRQASLLIMLQOQC
 LQTVGIRKPMILGAGNTSOSTHNTOKDOEYI JFRDRII SLVATSAVEGCLDQOC
 NVVATGILNDSINVOQARAGRAQSVSEIATGCSKRNRELIATNALLVEMKRAVA
 AVQKMDPEFRFAKIDLOQASLVKAAVAARHILQGGHILPEHVQALINQAVVGG
 SDLRVVECTHHVNNPNFSVYTTISUNGVVINKYFKDKKKGCTIRGSKCSVEGFMQI
 YKSVTLPLAKIGSMLETPRCKTIQKKSRVSPVPIVLIQDCTQSLSTSLID"
 BASE COUNT 591 a 594 c 655 g 485 t 1 others
 ORIGIN
 Alignment Scores:
 Pred. No.: 1.44e-83 Length: 2326
 Score: 1356.00 Matches: 299
 Percent Similarity: 59.69% Conservative: 129
 Best Local Similarity: 41.70% Mismatches: 237
 Query Match: 25.53% Indels: 52
 DB: 10 Gaps: 12
 US-09-515-363c-2 (1-1025) x AF316999 (1-2326)

QY 306 LeuGlnLeuAArgProTyrGlnMetGluValAlaGlnProAlaLeuGluGlyAsnLeu 325
 DB 2 ATGGAGCTCCGACCCCTACAGAGTGGAGATCTTACCTGCTGGAGGCAAGATATC 61
 QY 326 IleLeuGlyLeuProThyGlySerGlyLysThrArgValAlaValTyrTlaValAsp 345
 DB 62 ATTATATGGCTGCCCGGAGCTGCTGGCAAGACCGGCGAGCTGCTGTGATATCAAGAG 121
 QY 346 HisLeuAspLysLysLysAlaSerGlnProGlyLysValIleValLeuValAsnLys 365
 DB 122 CATCTAGAG-----ACGGTACACAGAGCGAGGTGGTGGTACTGCTGATACAG 169
 QY 366 ValLeuLeuValGluGlnLeuPheArgGlyGlnPheGlnProPheLeuLysTyrPyr 385
 DB 170 GTACACCTGAGTGGACAG-----CACGCCAGAGGATTACAGGCCCATGATGAATACAG 226
 QY 386 ArgValIleGlyLeuSerGlyAspPheGlnLeuLysIleSerPheProGlnValValLys 405
 DB 227 ACTGTACACACCTGAGTGGACAGACAGGATGCCATGCCAGCTGGCTGTGGCTGTGG 286
 QY 406 SerCysAspIleIleIleSerThrAlaGlnIleLeuGlnAsnSerLeuLeuAsnGln 425
 DB 287 ACCCAGACCTGCTGCTGCTGATGAGGACAGAGTGTACAGTGGCAGCTCAACATCTAG 346
 QY 426 AsnGlyGluAspAlaGlyValGlnLeuSerAspPheSerLeuIleIleAspGluGly 445
 DB 347 -----GAGGATGAGACACGCTGAGCTCAGAGATTCGCTGATGTGTGATGAGT 400
 QY 446 HisIleThrAsnLysGlnAlaValTyrAsnAsnIleMetAlaGlyTyrLeuMetGlnLys 465
 DB 401 CACACACCCACAGAGACACCGCTCTACACACCATCTGGACCCGCTACATGAAATAGAG 460
 QY 466 LeuLysAsnAsnArgLeuLysLysGluLysLysProValIleProLeuProIleLeu 485
 DB 461 CTGAG-----AAGCAGAG-----CCCTCCCTCCCTGCTG 493
 QY 486 GlyLeuThrAlaSerProGlyValGlyGlyValThrLysGlnAlaLysAlaGlnLys 505
 DB 494 GGTCTCACAGGCTCCCGGAGCAGTGGAGGCGCACCAAGGCGCATATGATCAG 553
 QY 506 IleLeuLysLeuCysAlaAsnLeuAspAlaPheThrIleLysThrValLysGluAsnLeu 525
 DB 554 ATCTTACAGCTTTGTCCCAATTTGGATACGTGCCATCATCATGCTGCCCAACATCTCTAG 613
 QY 526 AspGlnLeuLysAsnGlnIleGlnLeuProCysLysLysLysPheAlaIleAlaAspAlaThr 545
 DB 614 TCCGAGCTGCTGATGCATACCAACCAAGCCCTGCAGGAGTATGACCTGTGAAAGGCG 673
 QY 546 ArgGluAspProPheLysGlnLysLeuLeuGlnIleMetThrArgIleGlnIleLysCys 565
 DB 674 GCACGAGATCTTTGGGGGAGCTGTGATAAAGCTTATGAACCAATCCCACTCACTA 733
 QY 566 GlnMetSerPromet-----SerAspPheGlyThrGlnProTyrGlnGlnTyrAlaIleGln 584
 DB 734 GAGATGCTGCTGACTTAAACCAACATTTGAAACCCAGATGTATGACCAACCAATGATACAG 793
 QY 585 MetGluLysLysAlaAlaLysLysGlyAsnArgLysGluArgValLysAlaGlnLysLeu 604
 DB 794 TTGTCCAGAGATGCGCGAGAGCTGACTCCAGAACAGCGGGTATAGCCTGATTG 853
 QY 605 ArgLysTyrAsnGlnAlaLeuGlnIleAsnAspThrIleArgMetIleAspAlaLysThr 624
 DB 854 CGCGCTACATGATGCTGCTATTATCCAGCATACTGCTGCGCGGCTGCTTGCAG 913
 QY 625 HisLeuGlnTyrPheTyrAsnGlnGluLysAspLysLysPheAlaValIleGluAsp 644
 DB 914 ATGTTGCAAGATTTTACGACAGAGACGACCAACCAACAGATGCTGCTGCTGAA 973
 QY 645 SerAspGluGlyLysAspAlaLysLysCysAspLysLysGluAspLysLysLys 664
 DB 974 AGC----- 976
 QY 665 LysProLeuLysLeuAspGluThrAspArgPheLeuMetThrLeuPhePheGluAsn 684

Score: 1356.00 Matches: 299
Percent Similarity: 59.69% Conservative: 129
Best Local Similarity: 41.70% Mismatches: 217
Query Match: 25.53% Indels: 52
DB: 10 Gaps: 12

US-09-515-363c-2 (1-1025) x BC029209 (1-2336)

OY 306 LeuGlnLeuArgProTyrGlnMetGluValAlaGlnProAlaLeuGlnTyrLysAsnIle 325
DB 13 ATGACACTGCGACCCCTACAGAGTGGAGATCTTACCTGCTGTGGAGGCTAAGAAATATC 72

OY 326 IleIleCysLeuProThrGlySerGlyLysThrArgValAlaValIleValAlaLysAsp 345
DB 73 ATTATATGCTGCGCCACGGGCTGTGGAGACCCGCGCCGCTTTTCAACCAAGAG 132

OY 346 HisLeuAspLysLysLysLysAlaSerGluProGlyLysValIleValIleValAlaLys 365
DB 133 CATCTAGAG-----ACGGTAGACACAGCGCAAGCTGTGTCTACCTTCATTAAC 180

OY 366 ValLeuLeuValGluGlnLeuPheArgLysGluPheGlnProPheLeuLysThrTyr 385
DB 181 GTACACTGCTGTAGCCAG---CACGGCCAGAGATTCAGTGCATGTGATTAACACTGG 237

OY 386 ArgValIleGlyLeuSerGlyAspThrGlnLeuLysIleSerPheProGlnValValLys 405
DB 238 ACCGTACCAACCTGTAGTGGGACATGGATCCCGAGCTTGGCTTGGCTGTGG 297

OY 406 SerCysAspIleIleIleSerThrAlaGlnIleLeuGlnAsnSerLeuLysLeuGln 425
DB 298 AGCCAGACCTGCTCTCTCTGTACGGCAGAGTTTACATTCGCACTCAATAGCTCTAG 357

OY 426 AsnGlyGluAspAlaGlyValGlnLeuSerAspPheSerLeuIleIleLysAspGlnLys 445
DB 358 -----GAGGATGACAACTAGACTAGCTACAGAGAATTCCTCATGTGTGCGACGAGTCT 411

OY 446 HisHisThrAspLysGlnValValIleAsnAsnIleMetArgHisLysLeuMetGlnLys 465
DB 412 CACACACCCACCAAGACACCCCTTACCAACCATCTTACCGCGGTACCTTAACACGAG 471

OY 466 LeuLysAsnAsnArgLeuLysGluAsnLysProValIleProLeuProGlnIleLeu 485
DB 472 CTGGAAG-----AAGGCAAG-----CCCTCCCTCAGGCTCG 504

OY 486 GlyLeuThrAlaSerProGlyValGlyValAlaThrLysGlnAlaLysAlaGlnGlnHis 505
DB 505 GGCTCCACGCTCCGCCAGGCACTGGAGGGGCCACCAAGCTCCCAAGGGGCTATGATCAG 564

OY 506 IleLeuLysLeuCysAlaAsnLeuAspAlaPheThrIleLysThrValLysGlnLysLeu 525
DB 565 ATCTTACACTTGTGCGAATTTAGTACGTCACCATATGTCGCCCAAGAAATTCCTTAC 624

OY 526 AspGlnLeuLysAsnGlnIleGlnLysProCysLysLysPheAlaIleIleLysAspAlaThr 545
DB 625 TCCACACTGCTCATGATTAACCCGAAAGCTCTCAACACATATGACTCTCTCAAGCTCC 684

OY 546 ArgGluAspProPheLysGlnLysLeuGlnIleMetThrArgIleGlnThrTyrCys 565
DB 685 GCACACGATCTTTGGGAGCTTGATAAAAGCTTATGACCAAAATCCACCAACACACTA 744

OY 566 GlnMetSerProMet-----SerAspPheGlyThrGlnProTyrGlnIleValIleGln 584
DB 745 GAGATGCTGACTGACGAGCAACAAATTTGGACACCCAGCATATGACAGCAAGCACTGAC 804

OY 585 MetGluLysLysAlaLysLysGlyAsnArgLysGluArgValCysAlaGlnLysLeu 604
DB 805 TTCTGCAAGAGATGCGGCAAGGCTGACTCCAGAACAGTGGGTGTATGCTGCAATTTG 864

OY 605 ArgLysTyrAsnGlnAlaLeuGlnIleAsnAspThrIleArgMetIleValAlaTyrThr 624
DB 865 CGCGGCTACAAATGATCGGTATTTATCCAGTACTGTGTGCCCGGAGTCCCTTAC 924

OY 625 HisLeuGlnThrPheTyrAsnGlnLysAspLysLysPheAlaValIleGlnLysAsp 644

DB 925 ATGTTGCAGAGATTTTACAGACAGACACGCCACCAAAACAGCATGCTGCTGCTCA 984

OY 645 SerAspGlnGlyGlyAspAspGlnTyrTyrCysAspGlyAspGluAspGluAspLeuLys 664
DB 985 AGC----- 987

OY 665 LysProLeuLysLeuAspGlnThrAspArgPheLeuMetThrPhePheGluAsnAsn 684
DB 988 -----TGGCTGTGAAGCTGTTGATGACATAA 1017

OY 685 LysMetLeuLysArgLeuAlaGluAsnProGlnTyrGlnAsnGlnLysLeuThrLysLeu 704
DB 1018 AATGCTGGGCGACCTAGCAGAGCTGGGCTCT---GAGAACCCGCAACTTGGAGATGTG 1074

OY 705 ArgAsnThrIleMetGlnGlnTyrThrArgThrGlnLysLeuIleValIleIlePhe 724
DB 1075 GAAAGATCTTACTAAGCAGTTT---GGCAGTCTGTGCCACACTGGGCTATATCTC 1131

OY 725 ThrLysThrArgGlnSerAlaTyrAlaLeuSerGlnTyrIleThrGlnAsnGlnLysPhe 744
DB 1132 ACCAGACCCGTCAGACTGCTCTCTGCTGCTGCTTGGCAGCAGCTTGTCTA 1191

OY 745 AlaGluValGlyValLysAlaHisHisLeuIleGlyAlaGlyHisSerSerGluPheLys 764
DB 1192 CAGACTGTGGGCATCAGCCACAGATGCTGATCGGACGAGGAGAACACAGCCAGACACA 1251

OY 765 ProMetThrGlnAsnGlnLysGluValIleSerLysPheArgThrGlyLysIleAsn 784
DB 1252 CACATGACCCCAAAAGACACAGCAGAGTGTATCCAGAGTTCAGAGATGATCTTAC 1311

OY 785 LeuLeuIleAlaThrThrValAlaGlnGlnLysLeuAspIleLysGlnCysAsnIleVal 804
DB 1312 CTTCAGTGGCCACAAAGTGTGGCAGAGAGGCGCTGCATTCGCTCACTGCATATGTCG 1371

OY 805 IleArgTyrGlyLeuValThrAsnGlnIleAlaMetValGlnAlaArgGlyArgAlaArg 824
DB 1372 GTGGCTATGCGCTCTTACCAATGATGATTCACATGCTGCGACGGCGGCTGTGCTGA 1431

OY 825 AlaAspGlnSerThrThrValLeuValAlaHisSerGlySerGlyValIleGlnHisGln 844
DB 1432 GCTGTGACAGCTGTACTCTCTCTGCTACAGAGGCACTCGGAGATTAAGCTGAG 1491

OY 845 ThrValAsnAspPheArgGlnLysMetMetTyrLysAlaIleHisCysValGlnAsnMet 864
DB 1492 CTAAACCAATGAGGCTGTGAGCTGTGATGAGAAAGCTGTGGCTGTGACAAAGATG 1551

OY 865 LysProGlnGlnTyrAlaHisLysIleLeuGlnLeuGlnMetGlnSerIleMetGlnLys 884
DB 1552 GACCTGTAGAGTTCAGGCGCAAGATCGGAGACTTGCAGCAAGCATCTCTGTTAAGCG 1611

OY 885 LysMetLysThrLysArgAsnIleAlaLysHisTyrLysAsnAsnProSerLeuIleThr 904
DB 1612 GCAGACGCGGGCGCCATCGGAGATCCAGACAGGGCGAGTTCCTACCGGACACTGTGA 1671

OY 905 PheLeuCysLysAsnCysSerValLeuAlaCysSerGlyGluAspIleHisValIleGln 924
DB 1672 CTCTGCTCATCAATCATGATGATGCGCGTGGGCTACAGAGATCGCTGCGGAAAGTGG 1731

OY 925 LysMetHisHisValAsnMetThrProGlnLysGlnLysGlnLeuTyrIleValArgGlnAsn 944
DB 1732 GGCACCCACACGCTCAATGATCAACCCCAACTTCTCGCTACTATACCACTCCCAAC 1791

OY 945 Lys-----AlaLeuGlnLysLysCysAlaLysPyrGlnIleAsnGlyGlnIleCys--- 962
DB 1792 CCTGTGGCTATTACAAAGGCTCTTAAGGACTGAGACCTGGAAGAACCATAGTGTAGT 1851

OY 963 LysCysGlnGlnAlaTyrGlyThrMetMetValHisLysGlyLeuAspLeuProCysLeu 982
DB 1852 AACTGTGGAGGCTGTGGGCTTCCAGATGATCTACAAATGACGACTTGTCAATGCTC 1911

OY 983 LysIleArgAsnPheValValIlePheLysAsnAsnSerThrLysLysGlnTyrLysLys 1002
DB 1911 ----- 1911

BASE COUNT 591 a 737 c 783 g 502 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1,09e-82 Length: 2613
 Score: 1344.00 Matches: 291
 Percent Similarity: 59.55% Conserved: 136
 Best Local Similarity: 40.59% Mismatches: 238
 Query Match: 25.31% Indels: 52
 DB: Gaps: 12

US-09-515-363C-2 (1-1025) x AK021416 (1-2613)

QY 306 LeuGlnLeuArgProTyrGluMetGluValAlaGlnProAlaLeuGluLysAspIle 325
 DB 224 ATGGAGCTGGCTGCTACCAATGGAGGATGATACCTGCTGAGGCTCAAGAAATATC 283

QY 326 IleLeuLeuProThrGlySerGlyThrArgValAlaValTyrTAlaLysAsp 345
 DB 284 ATCATCTGGCTGCCAGCGGTGCCGGAAGACCGGCGCTGCTTATGTCACCAACGCG 343

QY 346 HisLeuAspLysLysLysLysAlaSerGluProGlyValIleValIleValLys 365
 DB 344 CACCTAGAG-----ACTGTGATGGAGCCCAAGGTGGTGTATATGCTCAACAGG 391

QY 366 ValLeuLeuValGluGlnLeuPheArgLysGluPheGlnProPheLeuLysTyr 385
 DB 392 GTGCACCTGGTACCCAG-----CATGCTGAGAGTTCAAGGCGATCTGGATGG 448

QY 386 ArgValIleGlyLeuSerGlyAspThrGlnLeuLysIleSerPheProGluValLys 405
 DB 449 ACCGTGACCAACCTGAGTGGGACCATGGGACCATGCTGCTGGCTGGCACTGGCCCGG 508

QY 406 SerCysAspIleIleIleSerThrAlaGlnIleLeuGluAsnSerLeuLeuLeu 425
 DB 509 TGGCATGACCTGCTATGTCACAGCAGGCTTCGCAATGTCACATGCACGCCCGAG 568

QY 426 AsnGlyAlaAlaGlyAlaGlnLeuSerAspPheSerLeuIleIleAspGluLys 445
 DB 569 -----GAGGAGGACGCTGAGCTGCTCTCTCTCTCTATGCTGATGATGATGCTG 622

QY 446 HisHisThrAsnLysGluAlaValTyrAsnAsnIleMetArgHisTyrLeuMetGln 465
 DB 623 CACCAACGACACAGGACACGCTGTACAACGTCATCAGCAGCTACCTACCACTAA 682

QY 466 LeuLysAsnAsnArgLeuLysLysGluAsnLysProValIleProLeuProGlnIle 485
 DB 683 CTCAGAGGCGCACAG-----CCGCTACCTCAGGCTGCTG 715

QY 486 GlyLeuThrAlaSerProGlyValGlyAlaThrLysGlnAlaLysGlnGluHis 505
 DB 716 GGTTCACAGCCTCCCGACGACTGGCGGGCTCCCAACCTGATGGGATATCAACAC 775

QY 506 IleLeuLysLeuCysAlaAsnLeuAspAlaPheThrIleLysThrValLysLysLeu 525
 DB 776 GTCCTGACCTCTGTGCCAACTTGGACAGCTGTGATGATGATGACCCCAAACTCTG 835

QY 526 AspGlnLeuLysAsnGlnIleGlnGluProCysLysLysPheAlaIleAlaAspAla 545
 DB 836 CCCCAGCTGACGACACACCAACAGCCTTCACCAACATACACCTCTCTACAGGCGC 895

QY 546 ArgGluAspProPheLysGluLysLeuGluIleMetThrArgIleGlnThrTyrLys 565
 DB 896 ACCAGAGATCCGTTGGGACTGCTGAAGAAGCTCATGACCAATTCATACACACCTG 955

QY 566 GlnMetSerProMetSer---AspPheGlyThrGlnProTyrGluGlnTyrAlaIleGln 584
 DB 956 GAGATCCGCTGAGCGCAAAATTTGGGACCAATATATGAGAGAGAGAGTGGAG 1015

QY 585 MetGluLysAlaAlaLysLysGlnAsnArgLysGluArgValCysAlaGluHisLeu 604
 DB 1016 CTGAGTAGGCTGCGCTTGGCTGGGCTTCAGAGAGCAACGATGATGCTCTCACTG 1075

QY 605 ArgLysTrpAsnGluAlaLeuGlnIleAsnAspThrIleArgMetIleAspAlaThr 624
 DB 1076 AGCGGCTACATGACGCCCTCTCATCATACACCGTCCGGCGGTGATGCTTGGCT 1135

QY 625 HisLeuGlnThrPheTyrAsnGluLysAspLysPheAlaValIleGluAsp 644
 DB 1136 GCGCTGACGATTTCTATACAGGAGAGCTCACTAAACCCAGATCTG----- 1186

QY 645 SerAspGluGlyLysAspGluTyrCysAspGlyAspGluAspAspPheLys 664
 DB 1187 -----TGT----- 1189

QY 665 LysProLeuLysLeuAspGluThrAspArgPheLeuMetThrLeuPheGluAsn 684
 DB 1190 -----GCCAGCGCGCTGCTGCGCTTCTGATGATGATGATGATGATGATGATG 1228

QY 685 LysMetLeuLysArgLeuAlaGluAsnProGluTyrGluAsnGluLysLeuThr 704
 DB 1229 AATGAGCTGGCCCACTGGCAACTGAGCCCA-----GAGATCCAAAACCTGGAATGCTG 1285

QY 705 ArgAsnThrIleMetGluGlnTyrThrArgThrGluSerAlaArgGlyTyrIlePhe 724
 DB 1286 GAAAGATCTCCAAAGCCAGCTGAGT---AGCTCTAACACCTCGGGGTATATCTTC 1342

QY 725 ThrLysThrArgLysSerAlaTyrAlaLeuSerGlnTyrPheThrGlnAsnGluLysPhe 744
 DB 1343 ACCGACCGCCGCAAGGCGACACTCTCTGCTGTGCTGCTGACGACCAAAATGCTG 1402

QY 745 AlaGluValGlyValLysAlaHisIleLeuIleGlyAlaGlnHisSerSerGluLys 764
 DB 1403 CAGACTGTGACATCCGCGCCAGCTACTGATGATGGCTGGGAGACAGACGCGATACAC 1462

QY 765 ProMetThrGlnAsnGluGlnLysGluValIleSerLysPheArgTyrHisLeu 784
 DB 1463 CACATGACCCAGAGGAGGACCAAGAACTGATCCAGAACTGATGATGATGATGATG 1522

QY 785 LeuLeuIleAlaThrThrValAlaGluGluLysLeuAspIleLysGluCysAsnIleVal 804
 DB 1523 CTTCGTGGGCGACAGAGTGGCGGAGAGGCGCTGACATCCCAATGCAATGCAATG 1582

QY 805 IleArgTyrGlyLeuValThrAsnGluIleAlaMetValGlnAlaArgAlaArg 824
 DB 1583 GTGCGTATGGGCTTGTACCAATGAAATCTGATGCGCAGGCGAGGCGGCTGCTG 1642

QY 825 AlaAspGluSerThrTyrValLeuValAlaHisSerGlySerGlyValIleGlnGlu 844
 DB 1643 GCGCATCAGAGTGTATACGCTTGTACCACTGAAGGTACCGGACCTGAACCGGAG 1702

QY 845 ThrValAsnAspPheArgLysLysMetMetTyrLysAlaIleHisCysValGlnAsnMet 864
 DB 1703 CTGATCAACGAGCGCTGAGACGCTAATGAGAGCGAGAGTGGCTGCTGCTGCAATAATG 1762

QY 865 LysProGluGluTyrAlaHisLysIleLeuGlnLeuGlnMetGlnSerIleMetGlnLys 884
 DB 1763 GACCGCGCGGAGTACGACGACGATCCGAGATCTGACGAGGACGCTTACCAACGCG 1822

QY 885 LysMetLysThrLysArgAsnIleAlaLysHisTyrLysAsnAsnProSerIleIleThr 904
 DB 1823 CGCGCCACGACCGCCAGCGGAGAACCGGACGAGCACTTCCAGTGGAGATATGTCAC 1882

QY 905 PheLeuCysLysAsnCysSerValLeuAlaCysSerGlyGluAspIleHisValIleGln 924
 DB 1883 CTACTCTGATCAACATGATGATGCTGTGGGCCATGGCAGACGACCTGGGAAATGAG 1942

QY 925 LysMetHisHisValAsnMetThrProGluPheLysGluLeuTyr---IleValArgGln 943
 DB 1943 GGCACCCACATGCAATGTAACCCCAACTCTCGAATCTATATGTCATATGATAT 2002

QY 944 AsnLysAlaLeuGlnLysLysCysAlaLysPyrGlnIleAsnGlyGluIleIleCysLys 963
 DB 2003 CCTGTGTCATCAACAAAGTCTTCAGAGCGAAGCCTGGGGGTGTCATATATGAGG 2062

QY 964 ---CysGlyGlnAlaTyrGlyThrMetMetValHisLysGlyLeuAspLeuProLysLeu 982

GenCore version 5.1.4.P5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 15:41:47 ; Search time 30 seconds
(without alignments)
1005.283 Million cell updates/sec

Title: US-09-515-363C-2
Perfect score: 5311
Sequence: 1 MSGNSTDENFRYLSCFRA.....LPITFPNDYSECCLFSPED 1025

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents-AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTCUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfl.es1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	194.5	3.7	2482	1	US-08-328-254-6 Sequence 6, Appl
2	191.5	3.6	3248	1	US-08-353-700-1 Sequence 1, Appl
3	191.5	3.6	3248	5	PCT-US95-16216-1 Sequence 1, Appl
4	188.5	3.5	1388	2	US-08-685-576-4 Sequence 4, Appl
5	186.5	3.5	1786	4	US-08-973-462-9 Sequence 5465, Ap
6	179	3.4	666	4	US-09-134-001C-5465 Sequence 5465, Ap
7	174	3.3	1388	4	US-08-685-576-1 Sequence 1, Appl
8	173	3.3	1211	4	US-09-134-001C-4820 Sequence 4820, Ap
9	171.5	3.2	976	4	US-09-104-324B-4 Sequence 1, Appl
10	171.5	3.2	2285	4	US-09-308-375-2 Sequence 2, Appl
11	167	3.1	3696	4	US-09-134-001C-5080 Sequence 5080, Ap
12	166.5	3.1	1886	4	US-08-938-105-3 Sequence 1, Appl
13	164	3.1	1354	3	US-08-685-871-2 Sequence 2, Appl
14	163	3.1	1939	4	US-09-310-187A-1 Sequence 3, Appl
15	159.5	3.0	956	4	US-09-134-001C-4452 Sequence 4452, Ap
16	158	3.0	781	2	US-08-675-631-3 Sequence 1, Appl
17	155	2.9	1261	4	US-09-208-742-4 Sequence 1, Appl
18	155	2.9	1261	4	US-09-332-295-2 Sequence 2, Appl
19	155	2.9	1261	4	US-09-332-295-2 Sequence 2, Appl
20	154.5	2.9	10182	4	US-09-709-979-2 Sequence 3159, Ap
21	152	2.9	407	3	US-08-989-370-6 Sequence 6, Appl
22	150	2.8	1234	4	US-09-592-054-8 Sequence 16, Appl
23	149	2.8	2329	3	US-08-755-587-16 Sequence 8, Appl
24	148.5	2.8	630	4	US-08-973-462-9 Sequence 9, Appl
25	148	2.8	407	3	US-08-989-370-5 Sequence 5, Appl
26	148	2.8	3418	2	US-08-639-501-2 Sequence 2, Appl
27	148	2.8	3418	3	US-09-044-946-2 Sequence 2, Appl

28	148	2.8	3418	3	US-09-044-908-2 Sequence 2, Appl
29	147.5	2.8	1461	2	US-08-993-228-10 Sequence 10, Appl
30	146.5	2.8	1098	4	US-08-923-992A-8 Sequence 8, Appl
31	146	2.7	989	4	US-08-213-419B-2 Sequence 2, Appl
32	146	2.7	989	4	US-08-213-419B-2 Sequence 2, Appl
33	145	2.7	2391	4	US-08-446-855A-2 Sequence 2, Appl
34	145	2.7	2391	4	US-09-150-741-2 Sequence 2, Appl
35	144	2.7	1388	4	US-09-572-191-2 Sequence 2, Appl
36	144	2.7	1388	4	US-09-723-262-2 Sequence 2, Appl
37	144	2.7	1388	4	US-09-723-219-2 Sequence 2, Appl
38	143	2.7	1232	4	US-09-592-054-2 Sequence 2, Appl
39	142.5	2.7	1312	2	US-08-687-080-51 Sequence 51, Appl
40	142	2.7	3418	2	US-08-603-753D-4 Sequence 4, Appl
41	142	2.7	3418	3	US-08-755-587-4 Sequence 44, Appl
42	142	2.7	3418	4	US-09-099-753-4 Sequence 4, Appl
43	142	2.7	3418	4	US-08-986-106-4 Sequence 4, Appl
44	141	2.7	1057	4	US-09-541-782-10 Sequence 10, Appl
45	141	2.7	1057	4	US-09-723-820-10 Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-08-328-254-6
Sequence 6, Application US/08328254
Patent No. 5710022
GENERAL INFORMATION:
APPLICANT: Zhu, Xueliang
APPLICANT: Lee, Wen-Hua
TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/328, 254
FILING DATE: 24-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/141, 239
FILING DATE: 22-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-CJ 1191
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2482 amino acids
TYPE: amino acid
TOPOLOGY: Linear
MOLECULE TYPE: protein
US-08-328-254-6
Query Match
Best Local Similarity 19.2%: Score 194.5; DB 1; Length 2482;
Matches 219; Conservative 186; Mismatches 473; Indels 265; Gaps 46;
QY 21 RYKMYIQVEPVL-----DYLTPLPAKVEQIQRTVATSG-----NNQAVELLIS 64
DB 1189 KIEACIELEKIVGELKENSDESKLEVFSCCHDELQVETSEGLNSDLQEMHAKSSRE 1248

QY 65 TLEKGVML--GWTREFEALRTGSPPLAARYMPELTLDPSPENAHILVLIQILNIO 122
 Db 1249 DIGDNVAKVNDKWKREFLD-----VENELSRIRSEKASIEH VALYLEADLE 1294
 QY 123 PTLVYDKL-LVRDYLDK-----CMEBELTJI-EDNRILAAENNGNE--S-VVELLKRIYQ 173
 Db 1295 VVOTELKLEKDNENKOKIVICLEELSVTSERNOLRGELDTMSKKTALIQOLSKRKE 1354
 QY 174 K-----ENMFSAFLNVLROT-----GNNELVDELTGSDCESNAELIENLS-VDGPOVEFQ 223
 Db 1355 KTOELSHQSECLHICQVAAEAEVEKTELLQTLIS-SDVSELLKDKRTHLOKIQLSKDSQ 1413
 QY 224 LLST-----VOPNLEKEVGMENNS-----SESSFADSSVSESJSTLAEGS---- 266
 Db 1414 ALSTLCGELENOIAOLNKKELLYKSESLOARLSSEDEKLVNSKALEVALVEKREFAL 1473
 QY 267 -VSCDELSGHSNMGSDSGTMSDSE-----ENVAARASPIPELOI KTYOMVAPQ 318
 Db 1474 RLSTOEEV-HOLRKGIEKLVRIEADKKOLHIAKLEKEREENDSLKI KVENLEKREIQ 1532
 QY 319 ALEGNIIICLPTGSGTRVAVYIAK-DHLDKKKASEPKVYIVLVNKVIVLEOLTKRKF 377
 Db 1533 MSENDELVIIDAKENKAEEVETLKTQIEEMARSLKVFELDVLTI RSEKENLTKOIQEKKG 1592
 QY 378 QPPLAKWYVIG-----LSGDTOLKISPEVYKSCDITLSTMOIENSLIEN----- 426
 Db 1593 Q--LSELIDLSSFKSLKEEKDEQAEIQIKEESK-----TAVEMLOQIKTLNEAVALC 1644
 QY 427 GEDAGVQLSDPSLIIDECHHTNKAEVYNNIRHYLMOKIKNNRLKKEKVIPIPLQI-- 484
 Db 1645 GDOELKATQSLDPIIEEFHOLRSTI-----EKLRARIEADKKOLCVLDOLKE 1694
 QY 485 -----LGLTASPGVGATKQAKAEHEILKI-----CANAD 514
 Db 1695 SEHHADLLKGRVENLERELEIARTNOEHAALEAKENSGEVETLKKIEGQSLRGELID 1754
 QY 515 APTIKVKNLDOLOKNOIQEPCKKFIAIADATREDPKKELLEINRIQTYVQMSPRSDG 574
 Db 1755 VVITREKENLNELOKEBERISELEIINSFENIIOEK-----EOKRVMEKESSTA 1807
 QY 575 TOPYEQMAIOMEKKAKKKNRERVCA--EHLRKYNEALQIN--CTIRHIIA-----YTHL 626
 Db 1808 MEMLOTOLKEINRVALINDQACAKKONLSQVECELEKALJLQGI DPAKNNYIYL 1867
 QY 627 EFTFYN-----EKKOKFAVIDSDSGDDEDEDDI KKPPLKLEDTD 673
 Db 1868 GSSVNGLIQEVERDGKOKLEKKDEISRLKNOI-----DQOEQVSKLSQVEGE 1915
 QY 674 RLMLTLPFENNMLKRLAENPEY-----ENELTKIKLNTIMFYTRPHI SARGIIFTK 726
 Db 1916 H--OLMKDONLELRNLVLEOKIOVLOSNNASLDDTLLEVLIQSSKNLENE--LELTK 1969
 QY 727 TROSAYALLSOMITENEKFAEYGVKAHHLIGASHSEFKPMONEJKEVI-KERTGCHNL 786
 Db 1970 MOKMSV-----EKVAKMTAKETELORENHEMAOKTAELOHELSGKNKLAGLLOL 2021
 QY 787 -----IATVAEBGLDICEKNIVIRGLVLTNEIAMVQANPARADESTYV 831
 Db 2022 LBEIKSKDQLKEITLLENSELKSLDCMHRDQVEKEGKREELIARYOLTHIAEKKHQAL 2081
 QY 832 LVAHSSGVILEHETVNDPFEKKMYKALHCVOMKRP-----EYAHKIFLOMOSIME 883
 Db 2082 LLDTNQYVEVEIOT--YEKELTSKE-ECSSQOKLEIDLKSKREFLNNSKATTO-ILE 2136
 QY 884 KKKMTK-----RNLAKHYKNNPSLITFLKNCOSYLACSGEDIVYIEK----- 925
 Db 2137 ELAKTKMDLKVYNOLKKNERAGOKMALLITSKQLE--EKKILQOKHISQLODAQOK 2193
 QY 926 -----MHIVNMPPEFKELYIVRENKA-----LOKKGADYQVINGELITKCGOAGMTM 971
 Db 2194 OKGTVMOTKVDLTELTEIKELKETLEEKTEKAEDEYLDKYS-LLISHKIKAKAKMLETQ 2252

QY 972 MVH 974
 Db 2253 VAH 2255
 RESULT 2
 US-08-353-700-1
 : Sequence 1, Application US/08353700
 : Patent No. 5599919
 : GENERAL INFORMATION:
 : APPLICANT: YEN, TIMOTHY J.
 : TITLE OF INVENTION: NUCLEIC ACID ENCODING A
 : TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
 : NUMBER OF SEQUENCES: 4
 : CORRESPONDENCE ADDRESS:
 : ADDRESS: DANN, DORFMAN, HERRELL AND SKILLMAN
 : STREET: 1601 MARKET STREET, SUITE 720
 : CITY: PHILADELPHIA
 : STATE: PA
 : COUNTRY: USA
 : ZIP: 19103-2307
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/353,700
 : FILING DATE: 09-DEC-1994
 : CLASSIFICATION: 435
 : ATTORNEY/AGENT INFORMATION:
 : NAME: REED, JANET E.
 : REGISTRATION NUMBER: 36,252
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (215) 563-4100
 : TELEFAX: (215) 563-4044
 : INFORMATION FOR SEQ ID NO: 1:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 3248 amino acids
 : TYPE: amino acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein
 : HYPOTHEetical: NO
 : ANTI-SENSE: NO
 : ORIGINAL SOURCE:
 : ORGANISM: HUMAN
 : US-08-353-700-1
 Query Match 3.68; Score 191.5; DB 1; Length 3248
 Best Local Similarity 19.28; Pred. No. 1.2e-06;
 Matches 221; Conservative 184; Mismatches 465; Indels 241; Gaps 47;

QY 21 RYKMTIQVEPV-----DYLTFLPAEYKQIORTVATG-----NMCAVHILLS 64
 Db 1917 KIEACIELEKELVGLKKNLSKLEVEFSCDHQELLQRETSGLNSDLEMAKSSRE 1976
 QY 65 TLEKGVML--GWTREFEALRTGSPPLAARYMPELTLDPSPENAHILVLIQILNIO 122
 Db 1977 DIGDNVAKVNDKWKREFLD-----VENELSRIRSEKASIEH-VALYLEADLE 2022
 QY 123 PTLVYDKL-LVRDYLDK-----CMEBELTJI-EDNRILAAENNGNE--SGVRHLIKRIYQ 173
 Db 2023 VVOTELKLEKDNENKOKIVICLEELSVTSERNOLRGELDTMSKKTALIQOLSKRKE 2082
 QY 174 K-----ENMFSAFLNVLROT-----GNNELVDELTGSDCESNAELIENLSQ-VIRTYVEEQ 223
 Db 2083 KTOELSHQSECLHICQVAAEAEVEKTELLQTLIS-SDVSELLKDKRTHLOKIQLSKDSQ 2141
 QY 224 LLST-----VOPNLEKEVGMENNS-----SESSFADSSVSESJSTLAEGS---- 266

```

Db 2142 ALSTKCELENOIAOLNKKELLYKESSELOARLESDEYKILNVSQKHLIAVKEGEPAL 2201
QY 267 -VSCLEDSIGHNSNMGSDGTMSDSE-----ENVARASPEPELOP RPYOMVAP 318
Db 2202 RLSTJOEEV-HOLRGIEKLRVRIEADKKOHLIAEKLERENDSLK KVENLEHLO 2260
QY 319 ALEGNIIICLPTSGKTRVAVYIAK-DHLDKKKKASEPGKVIYLVNKV LVEQLERKEF 377
Db 2261 MSENDELVIDAENKAEVETLKTQIEEMARSLKIFELDVLTRSEKINIKTOIOEKOG 2320
QY 378 QPFLKRWYVIG-----LSGDTOLKISPEYVKSODIISTAOILENSUNEN----- 426
Db 2321 Q--LSELDLSSFKSLBEKEQAEIOIKESK-----TAVEMLONOKILNEAVALC 2372
QY 427 GEDAGVLSDFSLIIDECHHTNKKAUVYNNIMRHYLMOKLNRLKKNKVIPLPOT-- 484
Db 2373 GQELMKKATGOSLDPRIEEOHILRSI-----ELTKARLEADKKQLCVLOJKE 2422
QY 485 -----LGLTASPGVGAATKQAKAEHILKT-----CANLD 514
Db 2423 SEHHADLLKGRVENLERELEIARTNOEHAALEENSKGEVETLKAKIEVY OSLRGLELD 2482
QY 515 APTIKTVKENDLQKNOIOEPCKKFAIADATREDPEKLEIMTHIOTVQMSPSDEG 574
Db 2483 VVIRSEKENLTELQEKQERISELIIINSFENILOEK----- 2521
QY 575 TOPYEMALOMEKKA-----KKGNKREKVCABH-----LAKTICELQIN-- 614
Db 2522 ---EOKYOMKESSTAMEMLOTOLKEIN--ERVAALINDQACAKELNISOVECLH 2575
QY 615 -DTRMIDAYTHLETFYNEKDKKFAVIEDSD-----EGGDEYC--TGEDEDLK 665
Db 2576 LEKAQLOGLDEKKNYIYLOSVKGLIOVEDGOKLKKDEISHLKN IODOLOLS 2635
QY 666 PKLDETFPLMTLFFENKMKRLAENPEY-----ENEKLTILKNTIMEQYRTES 718
Db 2636 KLSQVGEH---OLMKOENLELNLTVLEOKIOTVQSKNASLODTLEVY SSYKNLENE 2692
QY 719 ARCLIFTRQSAVYALSOMITENKFAEYGVANHL IGAGHSEFPFPMONOKYEYSAP 778
Db 2693 ---LELTAKDKMSFV-----EKVNMKTAKETELOREMHEMAQTAELCELSCEKNK 2741
QY 779 RTGKINL-----IATTVAEGLDIKECNIIVIRGLVYTN AMVQARGRA 823
Db 2742 IAGELOLLLEIKSSKQOLKELTLENSLKSJLCKMHKQVEKSVRGEAEYOLRLHE 2801
QY 824 RADESTYVIVANSGSVIEHETVNDPERKMYKAIHQVQNMKE-----EYAKKILB 875
Db 2802 AERKHQALLIDTNKOYEVEIOT---YREKITSKE-ECLSSOKLEIOLLSKEELNNSLK 2857
QY 876 LQMSIMEKKMKTK-----RIIAKHYNKNSPLITFLCKNSVLAOSGLIHVIEK-- 925
Db 2858 ATTO-LLEELKTKMDNLKYVNOJLKENREDAQKMLLIRSKOLE--EKEKILOKELS 2913
QY 926 -----MHVNMTPPEKELIYVRENKA-----LQKCAVOYNGEITCK 963
Db 2914 QLOAABEKOKTGVMDTKYDELTEIKELKETLEKTKREADFLKYCS LLSHKKLEK 2972
QY 964 CGGAMGTMAVH 974
Db 2973 AKEMLETOVAH 2983

```

```

RESULT 3
PCT-US95-16216-1
: Sequence 1, Application PC/TUS9516216
: GENERAL INFORMATION:
: APPLICANT: Yen, Timothy J.
: APPLICANT: Ratner, Jerome B.
: TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
: TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:

```

```

: ADDRESSEE: Dann, Dorfman, Herrell and Skillman
: STREET: 1601 Market Street Suite 720
: CITY: Philadelphia
: STATE: PA
: COUNTRY: USA
: ZIP: 19103-2307
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/16216
: FILING DATE:
: CLASSIFICATION:
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 08/353,700
: FILING DATE: 09-DEC-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Reed, Janet E.
: REGISTRATION NUMBER: 36,252
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 563-4100
: TELEFAX: (215) 563-4044
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3248 amino acids
: TYPE: amino acid
: STRANDEDNESS: not relevant
: TOPOLOGY: not relevant
: MOLECULE TYPE: protein
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: PCT-US95-16216-1

Query Match
Best Local Similarity 3.6%; Score 191.5; DB 5; Length 3248;
Matches 221; Conservative 184; Mismatches 465; Indels 291; Gaps 47;

: 21 RYKMTIQVEPVL-----DYTLPLPAVEKQIOQTVAATG-----NMQAVELLIS 64
: 1917 KIEACIELEKIEYGELEKENSDELSEKLEFSCDHOELLORVETSEGLNDEHMAHDSRE 1976
: 65 TLEKGYMHL--GMITREYFALKRTGSPPLAARNNELDLPSPSENAHDEYLC-LNLLQ 122
: 1977 DIGDVAKYNDSWKERFLD-----VENELSKIRSEKASIFH-EALVLEADFE 2022
: 123 PLYVDKL-LVRDVLK-----CMEELLTI-EDRRIRIAAENNGNE--SGVELKRIYQ 173
: 2023 VVOTELCKLEKNEKOKAYIVLEBELSVYTSERNOLGELDITMKSKTITADQLEKMKKE 2082
: 174 K----ENMFSAFLNVLROT-----GNNEIYQELTSGDSESNAELIENLSQ--VDGQVEEO 223
: 2083 KTOELSHQSECLHCLQVAEAEVKEKTELLQTL-SDVSELLKDKTHLOEKLOSLEKDSQ 2141
: 224 LSTT-----VOPNLEKEVGMENNS-----SESPASSVYSESDTSLAEGS----- 266
: 2142 ALSTKCELENOIAOLNKKELLYKESSELOARLESDEYKILNVSQKHLIAVKEGEPAL 2201
: 267 -VSCLEDSIGHNSNMGSDGTMSDSE-----ENVARASPEPELOP RPYOMVAP 318
: 2202 RLSTJOEEV-HOLRGIEKLRVRIEADKKOHLIAEKLERENDSLK KVENLEHLO 2260
: 319 ALEGNIIICLPTSGKTRVAVYIAK-DHLDKKKKASEPGKVIYLVNKV LVEQLERKEF 377
: 2261 MSENDELVIDAENKAEVETLKTQIEEMARSLKIFELDVLTRSEKINIKTOIOEKOG 2320
: 378 QPFLKRWYVIG-----LSGDTOLKISPEYVKSODIISTAOILENSUNEN----- 426
: 2321 Q--LSELDLSSFKSLBEKEQAEIOIKESK-----TAVEMLONOKILNEAVALC 2372
: 427 GEDAGVLSDFSLIIDECHHTNKKAUVYNNIMRHYLMOKLNRLKKNKVIPLPOT-- 484

```

Db 2373 GDOELKATDQSLDPFEEHOLRNSI-----EILRLNLEADKKQVCLQJUKK 2422
 QY 485 -----LGLTASPGVGATKQAKAHENHKL-----CANLD 514
 Db 2423 SEHHADLLKRGVENLERELEIARTNOEHALEAENSKGVETIKKIEGTQSLKLELD 2482
 QY 515 APTIKVKNELDLQKQIOEPCKKPAIADATREDPKKELLEMTRIQVTVMSPSDNG 574
 Db 2483 VVIREKSENLNLEOKEDERISELEIINSFENIIQEK----- 2521
 QY 575 TOPPEQMAIOMERKAA-----KCGNRKREVCSEH-----LRKYNALQIN--- 614
 Db 2522 ----EOKVQMKESSTAMBLQTOLEKLN--ERVAALINDQACKAKIUNISSVEGCE 2575
 QY 615 -DTIMIDAYTHLETFYNEKDKKFAVIEDSD-----KGDDEVC--TGDDEDDKK 665
 Db 2576 LEMAQLOLDLEAKKNYIYLOSQKGLIOVEDGKOKLEKDEISRLKNOIODOHOLVS 2635
 QY 666 PLKLDSTDFLMTLPFENKMKRLAENPEY-----INEXLTKLRNTIMEQYTRIEHS 718
 Db 2636 KLSQVEGEN---OLMKEONLELRLNLTVELOKIOVLQSNASIOBLEVIOSSYNLENI 2692
 QY 719 ARGIITKTROSAYALSONITENEKFAEYGVAKAHILIGASHSSEKPMPLNEOKEVISNF 778
 Db 2693 ----LELTAKDKMSFV-----EKYNNKTKMETELQRIHMEKQAKIIOBELSGEKKK 2741
 QY 779 KTKINLL-----IATVAEGLDKKCHIVIRGLVTVNIAWVOARSHFA 823
 Db 2742 LAGLELLOLLEIKSSKQDLKELELSELSKSLDCMHKIDJVEKGVKREIAEYOJRLHE 2801
 QY 824 RADESTYVVAHSGSVIEHETPNDRPKMNYKAIHGVNNMPE-----EYAKHILF 875
 Db 2802 AEKKHQAALLDNTKQYEVLEIOT---YREKLSKE--ECLSSQKIEIDLKSKSEELNLSK 2857
 QY 876 LQMSIMERKMTK-----RNIAKHYNNSPLITFLKNCVSALASDIDHVLTK 925
 Db 2858 ATTO-LLEFLKTKKMDNLKYVQKKNERAOQKMKLLKSKQOLE---KIKELIDKHEIS 2913
 QY 926 -----MHVNMTPFEKELYIVRENKA-----LOKCA VOINBETICK 963
 Db 2914 OLQAQEKOKGTVMQKVDLELITELIKELKLETKIADVLDKYCS-LIISHKLEK 2972
 QY 964 CGQAGTMMVH 974
 Db 2973 AKEMLETOVAH 2983
 RESULT 4
 US-08-685-576-4
 : Sequence 4, Application US/08685576
 : Patent No. 5906819
 : GENERAL INFORMATION:
 : APPLICANT: Kalbuchl, Kozo
 : APPLICANT: Iwamatsu, Akihiro
 : APPLICANT: Nakano, Takeshi
 : APPLICANT: Ito, Masaaki
 : APPLICANT: Takahashi, No. 5906819uaki
 : TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE
 : NUMBER OF SEQUENCES: 16
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Foley & Lardner
 : STREET: 3000 K Street, N.W., Suite 500
 : CITY: Washington
 : STATE: D.C.
 : COUNTRY: USA
 : ZIP: 20007-5109
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patent Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/685,576

: FILING DATE: 24-JUL-1996
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: JP 7-325129
 : FILING DATE: 20-NOV-1995
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: JP 8-17150
 : FILING DATE: 05-JAN-1996
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: JP 8-131206
 : FILING DATE: 26-APR-1996
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Bent, Stephen A.
 : REGISTRATION NUMBER: 29,768
 : REFERENCE/DOCKET NUMBER: 16887/843
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (202)672-5300
 : TELEFAX: (202)672-5399
 : TELEX: 904136
 : INFORMATION FOR SEQ ID NO: 4:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 138 amino acids
 : TYPE: amino acid
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein
 : US-08-685-576-4
 :
 : Query Match 3.58; Score 188.5; DB 2; Length 1348;
 : Best Local Similarity 18.24; Pred. No. 5.3e-07;
 : Matches 204; Conservative 195; Mismatches 399; Indels 321; Gaps 57;
 :
 QY 71 WHLGTREPEALRRTGSPLAARTYINPEL-TDLSPEFMAHDEYLOLNLILQI LVDKL 129
 Db 359 WH-W-----DNIRETAAPV-----PELSSDIDSSNFDIEDD---KQGVETFPPIPKA 402
 QY 130 LVROYLD-----KMEELITIEDRRIRIAAENNGESGVEBELKRIYQKEMFSAFLNLV 185
 Db 403 FVGNOLPFIIGTYIRENILLSDSP---SCRENDISIOSKNEESQIDQK-----LYTL 452
 QY 186 ROTGNNEI-VOELTSGDSCSESNAEIEENLSQVGDPOVEBOLSTVQPLKEKVMIMENNS 244
 Db 453 EEHLSENMAKKELEQKOKSVNTRLEKTAKE---ELEEEI---TLKRSVFALNOLEREK 505
 QY 245 SESSFAOSSVVSSEDTSL-----AEGSVSCIDESTL-----GHNSNMGSDS-GIMGSDSD 292
 Db 506 ALLOHKMAEYORKADHEADKRNLENVDVNSLKDQLEDLKKRNQNSQISTEFKVNIOGROLD 565
 QY 293 EENVAAARSPEPELQRLRYQMEVAQPA-----LECKNIILICLPTGSGKTVAAYIA 343
 Db 566 ETNALLRTESDTAARLRTQAESSKOIOOLESNNRDLODKNCL--LETAKLLEKEPINL 623
 QY 344 KDHLDKKKKASEPKVILVANKVLLVLEOLFPRKEFQPLKKVRYIGLSJDTOLKISPEEV 403
 Db 624 QSALESERDRDTHGSEII-----NDLQG-----RICGLERD----- 654
 QY 404 VKSCDIITIAQI-----LENSLNIENGEDAGVOIS-DESLIIT-----DECHII----- 447
 Db 655 LKNGLILAKVELERQIQSERFTDLK-EKSNMEIDMTYQKLVIOQSLEGEAEAKHAKTKA 713
 QY 448 --TKKEAVYNNIMRYLMQKLNKRLKRNKPVIRPQIQLLTASPGVGATKQAKAEH 505
 Db 714 RLADKNKIYESI-----EKAQSEAKMEKMKLE-----ERTLKQKVENL 753
 QY 506 IL---KLCANLDAFTIKVKNELDLQKQIOEPCKKPAIADATREDPKKELLEMTRIQ 562
 Db 754 LLEAKKRSILDC-----DLKQSQKQIKVELIKQ---KVLNED-VRNLTLEKIBEDQ 801
 QY 563 TYQWSPMSDFGTQPYEQMAIOMERKAAKKGKRRERYCAEHLKRYNEAL-----QI 613
 Db 802 KRC--LTQNDLKMQTOQVNTIKMSKQKQKQNNHLMKKKNALEKQNALEKREKQIADQGM 859
 QY 614 NDTIMIDAYTHLETFYV-----EKKOK-----KFAVIEDSD----- 646

Db 860 KELDOLAEFOYFLYKTOVRELKECEFKTKLGELOOKK/TL0DER:STA0LE1L 919
 QY 647 -----EGDDDEYCGDDED-----EDDKKPL-----KLTHTDRFLY LFENKML 687
 Db 920 TADSDOLARSTAEBOYSLPEKEKIMKEKMAHRKQELFKDATTISL-ETNRIL 978
 QY 688 KRLAENPEYENKLTIRNTIMEOYTRTEE---SARGILFTKROSAYA.SOMITEN--- 741
 Db 979 TSDVAILANEKEELNKKLDVOEQSLRLDELSMAAT-----KAOTKOLLTERTIK 1031
 QY 742 ----EKFAVGVKAHLLICAGHSSEFKPTOMOEKVISKFRJKNL----- 786
 Db 1032 TQAVNLALAI-MNRKEPVARGDIDVR-RKEKENRKLHMLKSRKELTLMIKYOKELN 1089
 QY 787 -LATYAEGLDICEKNITIRGYLTNETAMV0ARGBARA-----DESTVYLVHSGSGV 840
 Db 1090 BMOAQIAEE-----SOTRIEOLMTL0SKSDIDRLRSQLOALHT IDSSTSI -----GSGSP 1139
 QY 841 IEHETVNDPREKMM-YKAHGVONKPREYAKHLELOMOSIMEKKKIKRNIARHYN 898
 Db 1140 GDAEADDPRESLBEGLMLSPVRNNTKKRGVWKVIVSSKILFYDSGLKE-----QS 1194
 QY 899 NPSLITFLCKNSVLACSEDIHVLEKMHVNTPEFKEL-- YIVRENKALOKKADY 954
 Db 1195 NPMVVIDIDKLFHVRPTOTDVR-----RADAKEIPRIITLVANIGESKKEDEF 1244
 QY 955 QIN-GE-----IICKG0AGTMMVH----- 974
 Db 1245 PVEPVESKSNYIC0GHEPTPLTYHPTEC0ACMPMLHMFKIPAL0E0NCHINCKOH 1304
 QY 975 ----KGLDPLCKIRNFVVFKN-----NSTKKQYKKW 1004
 Db 1305 M0KKEELIAPC-KVYYDISTAKNLLILANSTEB0-QKW 1311

RESULT 5
 US-08-973-462-8
 : Sequence 8, Application US/08973462B
 : Patent No. 6191270
 : GENERAL INFORMATION:
 : APPLICANT: DUBUISSE, PIERRE
 : TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAIN: POLYPEPTIDE MOLECULES
 : FILE REFERENCE: 0660-0125-0 PCT
 : CURRENT APPLICATION NUMBER: US/08/973,462B
 : EARLIER APPLICATION NUMBER: PCT/FR96/00894
 : EARLIER FILING DATE: 1996-06-12
 : EARLIER APPLICATION NUMBER: FR 95/07007
 : EARLIER FILING DATE: 1995-06-13
 : NUMBER OF SEQ ID NOS: 29
 : SOFTWARE: PatentIn Ver. 2.0
 : SEQ ID NO 8
 : LENGTH: 1786
 : TYPE: PRT
 : ORGANISM: Artificial Sequence
 : FEATURE:
 : OTHER INFORMATION: Description of Artificial Sequence: Polypeptide
 US-08-973-462-8

Query Match 3.5%; Score 186.5; DB 4; Length 1786;
 Best local similarity 18.8%; Pred. No. 1.2e-06;
 Matches 215; Conservative 180; Mismatches 390; Indels 359; Gaps 50;
 QY 42 VKQDIOITVATSGNMAVELLSTLEKGVHLCMTREFEVALKRTSSPLAKRYMNFELID 101
 Db 323 VESVAENVEESVAENVEEIVAPTEIV-----APVLEIVASVVE 365
 QY 102 LPSPEFMADEVLQ-----LNLILQ 123
 Db 366 SVAPSVESVEENVEESVAENVEESVAENVEESVAENVEESVAENVEESVAENVEEIVAP 425
 QY 124 TL-----VDKLLVVDLCKMEELITTEDHNRITAAENNNESGVRLKRI0KEN 176

Db 426 TVEEIVAPTEIVAPSVESVAPSVESVEENVEESVAEN--VEESVAENVEEIVAPSV 483
 QY 177 WFSAPLAVIROTGNNEIYOEILTGSCSSMAETENLSOYDOP0VE0LLST-- 227
 Db 484 EESVAENVEESVAEN--VEESVAENVEESVA--ENVEEIVAPTEIVAPTEIVAPSV 539
 QY 228 --TTPNLEKEVGMENNSSESSP0DSSVSP0TSL0GVS0CDESLGNSN0D0SG 285
 Db 540 VESVAPSVESV--EENVEE-----SV0ENVEESVAEN--VEESVAENV----- 579
 QY 286 TMSDSDENVAARASPEPELOLREY0EVA0PALEGNNIITCPT--GSKTRVAVVIA 343
 Db 580 ---EESVAENVEEIVAPTEIVAPTEIVAPSV--VESVAPSVESVEENVEESVA 632
 QY 344 KDHLCKKKASPEGVYIYLVNKKVLLYB0LFKREPOPLKMYRVLGSLG0LKI0PEV 403
 Db 633 E---NVEESVAENVEESVAEN--VEEIVAPTEIV-----APVLEIVAPSV 675
 QY 404 VKSC-----DIIT0AILENSILNENG0DAG0L0D0SLI----- 440
 Db 676 VESVAPSVESVEENVEESVAENVEESV--AENVEESVAENVEESVAENVEEIVAPTEE 733
 QY 441 -----IIDECHHTNKEAVYNNI-----MRHYLMOKLKNRKLKKNRPYIP 480
 Db 734 IYAPTEIVAPSVESVAPSVESVEENVEESVAENVEESVAENVEESVAENVEESVAP 793
 QY 481 -LP0ILGLTASGVGATK0AKAEHILKLCANL-----D0F--TIKTYKE 523
 Db 794 TVEEIVAPSVESVAPSVESVAENVATNLSNLLG0LET0EIDISILNIEEYKE 853
 QY 524 N-----LDLKM0D0EPCKKFA-IADATRE-----DPKEXKLEI-----MRT0TY 564
 Db 854 NVYTTILENVEETTESVTTTSNILEEIOENTITDIT0EKL0EHL0ENVL0S0AENT0SE 913
 QY 565 COMSPSPDFT0PY0E0M0I-----0MEKKAKK-----GNRKERV-----C0EHLRK 606
 Db 914 EEKKEIV0IEVEEKEVATLIT0EY0AEK0SANTIT0EFL0EEN0AENS0ENVAENLEK 973
 QY 607 YNEALQINDT0MIDATYHL--ETFYNEBK0KKFVIED0SG0DDEYCGD----- 656
 Db 974 LNEIY-FNIVLDKVEEIVEIS0ESLENMEMD0KAFSEIFDVKY0I0ENLITGM-FSIFTS 1032
 QY 657 ----DEDEDLCKPLKIDT0RFLMTLFPENKMLKRLAENPEYENE-KLTKL0VTIMEQ 711
 Db 1033 IYI0SEKAYDL-----NENYVSSILDNIENKKEGLKLNINI----- 1069
 QY 712 YTRTESANGILFTKROSAYALS0MITENEKFAVGVKA--HHLIG-----AGHSSEF-- 763
 Db 1070 -----SSTGV0EYVTEH-----VE0NYV0DVDPAMK0D0FLGILN0AGG-KEMFF 1115
 QY 764 -----KPT0NE0K0EVI0SKRTGKINLITATVAE0ELIK0ECNI 803
 Db 1116 NLEDVF0KSESIVITVEEILDEY0K0EVEKEIY0IIEEM0EEN--IYDVLEEEKELC-ID-- 1170
 QY 804 VTRGYLTNETIAMV0ARGBARADESTYVYLVHSGSGVIEHETVNDPREKMMYKIHCVON 863
 Db 1171 -----KMLDA-----VEESIEI-----SSDSKERTESIK0-KEK0VSI VEE0V0D 1209
 QY 864 MKPEYAKHLELOMOSIMEKKK-----TKRNIARHYNKNPSL 902
 Db 1210 N0MDESEV0EVL0E--KNMEELIMKDAVEINDITSKLIE0T0GLNEVADLIDIMKLMEL 1267
 QY 903 ITFLCKNSVLACSGEDT--HYLEKMHVNMNT--PEFKELYIRENKAL0KCATY0INGE 959
 Db 1268 EKALS0DSKEIIDAKD0TLEKVIIEEH0ITTLDEVL0ELKDV0EDK1--EKVSI-KDEE 1325
 QY 960 IICK 963
 Db 1326 DILK 1329

RESULT 6

US-09-134-001C-5465
 : Sequence 5465: Application US/09134001C
 : Patent No. 6380370
 : GENERAL INFORMATION:
 : APPLICANT: Lynn Doucelte-Stamm et al
 : TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 : TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 : FILE REFERENCE: GTC-007
 : CURRENT APPLICATION NUMBER: US/09/134,001C
 : PRIOR FILING DATE: 1998-08-13
 : PRIOR APPLICATION NUMBER: US 60/064,964
 : PRIOR FILING DATE: 1997-11-08
 : PRIOR APPLICATION NUMBER: US 60/055,779
 : PRIOR FILING DATE: 1997-08-14
 : NUMBER OF SEQ ID NOS: 5674
 : SEQ ID NO 5465
 : LENGTH: 666
 : TYPE: PRT
 : ORGANISM: Staphylococcus epidermidis
 : US-09-134-001C-5465

Query Match 3.4%; Score 179; DB 4; Length 666;
 Best Local Similarity 19.5%; Pred. No. 1,1e-06;
 Matches 137; Conservative 114; Mismatches 230; Indels 220; Gaps 35;

QY 321 ECKNIIICL-PTGSGKTAAVYIANDHDKKKKASEPGK-VIVLVKKVLYVQLEFR--KE 376
 DB 37 ECKRHOTLLGATGCTGKTFMSHWIK-----EVSKPTLLIAHNKRLAGOLVSEFE 86
 QY 377 POPPLKKMYRVIGSGDDOLKISPEV-VKSCDIIISTAOJLNSLNLNFNEDGAYOLS 435
 DB 87 FPEENVEYFV-----SYDYDYPAYVSTDTFEKDCASINDEIDOLHSATSSLEPR 140
 QY 436 DESLIT--IDECHTNKEAVYNNIM-----RHYLMQKL-----KNN----- 469
 DB 141 DQVITIASVSCIYGLGNPREYKNLYSVAVGCMEMSESLKRLVQVYSHNIDRGRTF 200
 QY 470 RLKKNKRPVPLPOLIGLTASPGVGATQAKAEHILKLCANLDAF--T:KTVENILQ 527
 DB 201 RVRGDVEYFV-----ASREF--MCTRVFEFGDF:DRIRF-VAY 236
 QY 528 LKNOIEPCKKPAIDA-----TREDFK-----EKILE----- 556
 DB 237 LGEVIREHEHTTTPASFPVTRREMKVATIRIEKELEBERKEJRDENKILAEORLEPR 296
 QY 557 -----IMRIQTYCOMSPMSDPTQPY-----DQMAIOMK-----KA 589
 DB 297 TYNDEMMHREMGFGSGIENYSVHLRLPLGSPYTLDDYFGDDMLVMIDH:SVTLPLQIK 356
 QY 590 AKKGNR-KERVCAEHLKRNKALQINDTIRNIDATHTLETYNKEDKKAVAVIEDSDHG 648
 DB 357 MYNGDRARQVLLDHGFRPLSALD-NRPLK-----FREF--EKKTKOIVYV--SATP 403
 QY 649 GDEYCDGDEDEDLKKPLDDETRPLMTLFFENKMKRLAENPEVYN:KTLKLRNT 708
 DB 404 G-----PLELHTDENVQIIRPTGLLPKIDVRPT-ENK-----IDDL 442
 QY 709 MEOYRTTESARGIIFTKTROSAYALSOMITENKFAEYGVAAHHLIGASHSEKRPMTQ 768
 DB 443 SEIODRVDDEKLVLTLLTKKMSDELITYM-----KEAGIVANVL-----HSEIKTL- 489
 QY 769 NQKEEYISFRFGKINLLIATVAEGLDIKCNIVY-----RYGLVNF:AMVQARGR- 822
 DB 490 -EKIEITRDLRMGTDAIVGINLREGIDIPESLVVILDAQKFGFLKRSLSLTIOTIGRA 548
 QY 823 ARADESTYVL-----VAHSGSGVLEHETVNFPRKMMYKAH 859
 DB 549 ARNDGEVIMYADKTDSDMOYVIDETORRRELOIAHNKEHGIIPTIN-----KKIH 600
 QY 860 CVQNM-----KPEYAHKILELOMOSIMEKKMKTKRNIAKH 897
 DB 601 DIVISATVESDETNOOQTELPKMKTKKEROKTIENTEKEM 641

RESULT 7
 US-08-685-576-1
 : Sequence 1, Application US/08685576
 : Patent No. 5906819
 : GENERAL INFORMATION:
 : APPLICANT: Kalbuchi, Kojo
 : APPLICANT: Iwamatsu, Akihito
 : APPLICANT: Nakano, Takeshi
 : APPLICANT: Ito, Masaki
 : APPLICANT: Takahashi, No. 5906819aaki
 : TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE
 : NUMBER OF SEQUENCES: 16
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Foley & Lardner
 : STREET: 3000 K Street, N.W., Suite 500
 : CITY: Washington
 : STATE: D.C.
 : COUNTRY: USA
 : ZIP: 20007-5109
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/685,576
 : FILING DATE: 24-JUL-1996
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: JP 7-325129
 : FILING DATE: 20-NOV-1995
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: JP 8-17150
 : FILING DATE: 05-JAN-1996
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: JP 8-131206
 : FILING DATE: 26-APR-1996
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Beal, Stephen A.
 : REGISTRATION NUMBER: 29,768
 : REFERENCE/DOCKET NUMBER: 16887/843
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (202)672-5300
 : TELEFAX: (202)672-5399
 : TELEAX: 904136
 : INFORMATION FOR SEQ ID NO: 1:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 1388 amino acids
 : TYPE: amino acid
 : TOPOLOGY: linear
 : MOLECULE TYPE: protein
 : US-08-685-576-1

Query Match 3.3%; Score 174; DB 2; Length 1388;
 Best Local Similarity 18.5%; Pred. No. 9.5e-06;
 Matches 207; Conservative 186; Mismatches 390; Indels 334; Gaps 58;

QY 83 LRTSPPLAARMPEL-TDLSPSFENAHDEYQIIMLQPTIV-----D 127
 DB 364 IRETAAPV-----PELSSDIDSSNFDDIEDKDDVEFFIPKAFVGNOLPFIQTYRE 418
 QY 128 KLVADVLDKMEBELLTIEDNRNRIAAENNGNESGVRELLKRIYQENMFSAF:IVILQ 187
 DB 419 NLLISD-SPSCKEND-SIQSRK-----NEES-DEIOKKLYTLEHSLTEIO---- 461
 QY 188 TGNNEVLVELGSDCSSENAEIEMLSQYDGPQVEQOLSTVQNRNLEKRYMGKNNSSS 247
 DB 462 -AKELEOK-----CKSVNTRLEKVAK-----ELEEEI---TLRRVVESTLLQLEKRAL 508
 QY 248 SFADSVVSESQTSU-----AEGSVCLDESL-----GHSNNGSOS-CTMG:SNISDEN 295
 DB 509 QHKANEYORKADHEADKRNLENDVNSLKQLEDLKRNQNSOISTEKVYNQLOLQDFTN 568

```

QY 296 VARASPEPELOLPYOMVQOAP-----LEGNIIITLPGSGKIVAVYIANDH 346
DB 569 ALLRTESDPAARLKTQOAESSQIOLOESNNRDLQDK-----CL-----LETKAL 615
QY 347 LDKKKKASPEGVIVLVNKLVEOLFPRKEQPLFKKW RVIGI SGTQI KIS-PEPVK 405
DB 616 LEKE-----FINLSVLESEBRDRTHGSEII-----NDQITISGLEEDVK 656
QY 406 SCDDIITTAQI-----LENSLNLNGEDAGVOLS-DFSIITII-----D-THH-----447
DB 657 NCKIILAKLEKEKROLOEFDTLEK-EKNMEIDMTYQLKVLQSLSEOF EHKATKAL 715
QY 448 TKKAVYNNIMHMYLMOKKNNRKLKKENRPVLPOLILTASVGVGAI AKAEENIL 507
DB 716 AKNNKTYEST-----EKKSEAKMEKKLSE-----ERIKKOVENLIL 755
QY 508 ---KLCANIDAFITIKVENLDOLKNQLOEPCKFAIADATREPEKKELEIMTRIQTY 564
DB 756 EAKKSGSLDC-----DLKSGOOKINELLQO---KIVLNEO-VRNILIKIEDETOKR 803
QY 565 COMSPSDFGTOPYOMALOMKKAAGKNNKREVCALHRYNEAL-----QIND 615
DB 804 C-LTONDLKMQTOOVNLTLMSEKOLKORNNHLLKMKSLKONELRKI HODADGOMKE 861
QY 616 TIRMDATYHLETFYN-----EKKDK-----KFAVIEDSD-----646
DB 862 LODQLAEQVSTLYTQVRELKEBCESEKTLCKELQOQKOELDDBRDS-AOLEITLTK 921
QY 647 -----EGDDEYCDGDED-----EDDLKPL-----KLDETPELMTLPENNMKMR 689
DB 922 AUSEQIARSIAEBOYSDELEKIMKELEIKENMARIKOLTEKINATIASI EETNITLIS 980
QY 690 LBNPEYENKELTKLNTIMEOTRTEE--SARGITFTKQSAVALSILTEN-----741
DB 981 DVANLANEKEELNNKLEAOEOLSRKDELEISAAAI-----KAOFEKILTEETLTKQ 1033
QY 742 --EKFAEVNKAHNLGACHSSEFPKPTONEKEVYSKTRTKINL-----I 787
DB 1034 ANKLALET-MNKRERPKRNDIDVR-RKKEENKILMEIKSEREKITQOY-KYOKELNEM 1091
QY 788 ATTVAEGDIEKCNIVIRYGLVTNEIAMVQARGARA-----CESTUYIVANGSGVIE 842
DB 1092 QAOIABE-----SQIRLEIOMTLDKSDIEQLSQAQALHIGLSSSI-----GSGPDD 1141
QY 843 HETVNDPFRKMM--YKATCYONMKREYANHLLEOKMSIMEKKKTK--NIAKATKNNP 900
DB 1142 TEADDFPESRLEGWLSLVVRNNTKFGVVKYVIVSSKILFYUSEQDK-----QSNP 1196
QY 901 SLTFELCKMSYLAGSGEDIHYTEKHNHNMTPREKEL---YIVRENMTOJKKADYOL 956
DB 1197 YMLDIDKLFHVRVPTQDYU-----RADAKEIIRITQIYANEISKEQEPFV 1246
QY 957 N-GE-----IICKGQAGMTMVI-----974
DB 1247 ERVGESENITGHGHEFITLVHFTNCEAMKPLNMKRPVIAICRR-IKCHKCHMD 1306
QY 975 --KGLDPLCLIRNFYVVERKN-----NSTKKOYKAV 1304
DB 1307 KKEEIIAPC-KVYUDDISSAKNILLLANSTEEQ OKWV 1341

```

```

RESULT 8
US-09-134-001C-4820
; Sequence 4820, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCE RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13

```

```

; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; LENGTH: 1211
; TYPE: PRF
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4820
Query Match 3.3%; Score 173; DB 4; Length 1211;
Best Local Similarity 18.8%; Pred. No. 9,36-06;
Matches 210; Conservative 168; Mismatches 343; Indels 48; Gaps 52;
QY 44 EOIQRTVATSGNMAVELLSTLEKGVHLCMTREVEALRRTGSPPLAARYNN-ITDLP 103
DB 199 ESIOKLDHTEENLNVEDILYDLEGR-----VEPLKEEA-----233
QY 104 SPSFENADEVLOLNLQPTLVKLVYRDVLDKCMEEELITIEDNRRIA--ANNGNE 161
DB 234 -----IAKERYKOLSKEMEQS--DVIVTVSDIDHYETDMQRUDERLNHLSQA-KBQO 285
QY 162 SGVRELLKRI-----YOKENM-----FSAPLNVLRQNNELVSELTS 200
DB 286 AQINOLLKRYKGRQONDYDIKELNYELVATENTEOGLNVLNLEKKNQ-----337
QY 201 DCSSENA--EIEML-SQVDPQVE-----BOLLS--TTVOPNLEKRYWGMEN-----242
DB 338 --SETNARYEEELDLWLESQDISIKNEKAQONKELADLANKKOKOLNKEVQELESI YISDE 395
QY 243 -----NSESSEFSDSSVVS-----ESDLSLAGSVCSDSGSHSNM-SDSGT 286
DB 396 OHDELEIKNSYITLMSQSYVNNNDIRPLEHTTENENAKSRSDSLVEAFNC-KDIOQ 455
QY 287 MGSDEENVARASPEPELOLPYOMVQAPALEGKNIILCLPTGSKTRVAVIANDH 346
DB 456 NTQOKQEKVSSKGMK-KVEQNIQOLE--QOLDSKRL-----SEVENKIVQAVRY 505
QY 347 LDKKKK-----ASEGVKIVLVNKLVEOLFPRKEQPLFKKWIVVIGLSQDQ-KISFP 401
DB 506 MEKLSRISLATOEDDYTFPFGVKNILKAKDKELR-----GIRGAVAEVIVNP 555
QY 402 -EVVSCDIIITAOILLESLL-NLENGEDA-----GYOLDSFSLIIOECHIT-----448
DB 556 SMTQATIEALCAS--LOHVIYDNKDKROAIQYLKQKIGRATPLPLNVLOPR-VVADI 613
QY 449 -----NKEAVYNNIMHYL-----MOKIKN-NRLKKNRPVLPOL 484
DB 614 KQVANGSGOFINIASDAINVSAKYONTIENLNGTIIYENLKHANELARATRYHRIYTL 673
QY 485 ICLTASPG--VGATKOAKA-----EHLIKCANLDAFTIKT-----520
DB 674 ESDVNPFGSMTGGGARFKTSILSOKDELSTBRNLEDOYOROTAFEROFECKI-JAEDL 733
QY 521 -----VRE-----NLDOLEKNQ--IOPCKKPAIA--PATRECF-KEKLL 555
DB 734 SEQYFASQOYNNLKEQVHNHLELDRLCTOEAHLKHEHEBEFEKKNQYQCK-KETLK 793
QY 556 E-----IMTRIQTCOMSPSDFGTOPYOMALOMKKAAGKNN-ERYCA 601
DB 794 EKONHLEIQOOLKOLFSDIERYTQLSKKGKASTHOTQO--QLHQKSDLAUV-ERIKS 850
QY 602 -----EHLKRY-----NEALQINDITIRMI-----DAYHLETFYDEK-----634
DB 851 OKOVERELDKOLSDERQKTEVNERIKILNSDEMKGKDAFEKLOIOQOENVR--NLNQO 910
QY 635 -----DKFAVIEDSDDEGGDEYCGDDEDDIKKPLKDE-----TDR 674
DB 911 LSEIKQKRLMEKIEINSQLOKCHODILSIENHYODIKAKOSKIDVI-INHAHHLMDT 970
QY 675 FLMTLFFENNMKILKLAENPEYENE-----KLTKLR-----NTIM 709

```

DB 971 YOLTV-----EYDSDETINLKKKYLITMSIDELCPVNI NATEOFEELN 1020
 QY 710 EGYRTTESARGIIFTKROSAYALSOWITENKFAEYGVK AHILIGA-HISSEKPMTO 768
 DB 1021 ERYTFINEOR-----TDLREAKETLEQIITHENDKEVEGFKTTFAVQJPH TTVERKULFG 1075
 QY 769 NQKEY-----ISKFRTRK-----INLLTA-----788
 DB 1076 GGQAEILRTEDDYLSAGVDIIVOPPKKIQLHLSLSCGRALSAIALPLA LKVSAPFEV 1135
 QY 789 -TTVAEGEDIKECNIVIRGLVTEIAYVQARABAEFTYVIVAH-----835
 DB 1136 IDEVEALMD--EAN-VIRAYQYLNEL-----S-ETOPFIVITHKKGTMEERONLY 1182
 QY 836 ----SGSGVIEHETVNDFREKKMYKAIHCVQNNKPEEVA 870
 DB 1183 GVTMOESGVSKIVSVN-----LNTIDEVMEKEOA 1211

RESULT 9
 US-09-104-324B-4
 : Sequence 4, Application US/09104324B
 : Patent No. 6232460
 : GENERAL INFORMATION:
 : APPLICANT: T rect, Ozlem, Sahin, Ugur, Piremusistuh, v chael
 : TITLE OF INVENTION: Methods For Diagnosis And Treating Cancers,
 : TITLE OF INVENTION: And Methods For Identifying Pathogenic Markers In A Sample Of
 : TITLE OF INVENTION: NO. 6232460mal Cells
 : NUMBER OF SEQUENCES: 4
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Fulbright & Jaworski LLP
 : STREET: 666 Fifth Avenue
 : CITY: New York City
 : STATE: New York
 : ZIP: 10103
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage
 : COMPUTER: IBM
 : OPERATING SYSTEM: PC-DOS
 : SOFTWARE: Wordperfect
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/09/104,324B
 : FILING DATE: 25-June-1998
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: 08/892,702
 : FILING DATE: 15-July-1997
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Hanson, No. 6232460man D.
 : REGISTRATION NUMBER: 30,946
 : REFERENCE/DOCKET NUMBER: LUD 5491
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (212) 318-3000
 : TELEFAX: (212) 752-5958
 : INFORMATION FOR SEQ ID NO: 4:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 976 amino acids
 : TYPE: amino acid
 : TOPOLOGY: linear
 : US-09-104-324B-4

Query Match 3.2% Score 171.5; DB 4, Length 976;
 Best Local Similarity 20.5%; Pred. No. 8.9e-06;
 Matches 182; Conservative 132; Mismatches 333; Indels 239; Gaps 41;
 QY 136 DKCMEELITIEDNRIRIAAENGNGSGVRELKRTVQKKNH SAEFLVAVLQYTGNNEL-V 194
 DB 39 NCTEDDL-----EFPRAKTNLSKNGENIDSDPALQKN---FLPVLYQVNGNSDCHY 87
 QY 195 QR-LTGSDCSESNAEIEMLSOV-----DGPVEEQLISTTVQVULREKVGOMENNSSES 247
 DB 88 QEGKLDSOLENS-----EGLSRVFSKLYKEAEKIKKKKYSTVA -----ELRKQES 132

QY 248 SFADSVVSES-----DTSLAEGSVSC-LDESLGHSNMGSDSGTMSDSDEI NVAAE- 299
 DB 133 KLOENRKILIEAQRKALIQELOGNEKVSILKEEIGEN-----KDLIKE NNATRH 181
 QY 300 -----ASPEEIDLRTQYMEVAPALAEKNNIILICLPTGSGKRTVAAYIAHDLCK 349
 DB 162 LCNILKETCANSAEKTKYEEYERETROYMDLNNIEMKITAFOELNVAEANSLEHAF 241
 QY 350 KKKASEPKVILVAVKVLVLEOLFKEFOPFLKKWYVIGLSCTOLKISPEVAVKSCDI 409
 DB 242 KLKED-----YKIOHLOFYRKEL-----NKKENQVSL-----L 271
 QY 410 ITSTAQILNLSLN-----LENGEDAGVOLSDFSLI-----IIDECHITKEA -YVNN 456
 DB 272 LIQITE--KENMKKDLTFLLESRDKNVNOLEBKTKLOSENLSKOSTIEKOHHLTKEL FDIKVS 330
 QY 457 IMRHYLMOKLNNRLKNNKRPVIRPOLIGLTASGVGATKQAKAEHLIKT-VANDAF 516
 DB 331 LORSYSTOKALEEDLOIATKTIQOLTE-----BKEQOMESNKAHAAH--SF 375
 QY 517 T-----KTVKENLDOLNKOIOPCKKFA-IADATREDPKIKILEI 557
 DB 376 VYTERETTVCLEBELLRTQGRLEKNEQOLKILTMELOKSSELEPMKILTNKKIVELEE 435
 QY 558 MTRIOTYQMSPMDSGTOPYEQMAIOMEKKAAR-KGNRRKRVCAEHLIKYNEALJMDT 616
 DB 436 LKV-----LCKEETLAYEN--KQEKIAEELKGTQEDLLOLQARE-----KEYIDL 481
 QY 617 IRMIDAVTHLETFYNEE-KQKPAVIEDSDGDDGDDG-----DEDRILKKKPL 667
 DB 482 EIOLAITTSQYTSKEVKDKLTLENEKLNTELTSCHNLSLENKELIOTET-SMTELEL 541
 QY 668 KLDETRFLMTLFFENNK-----MLK-----RLAENPEYENKLTILKRNIMEQY 712
 DB 542 KNOQGD-----INNKKQOEERMLKQENQETETQJRLNELLEVVBELKQKHQVYCKL 594
 QY 713 TRTESASAGT-----ITKTRQSAVALSOWITENK--FAEYGVAAHHI IGAGHS 760
 DB 595 DKSEECNNLRKOVENKKNYIEILOENKALKKGTAESKOLNYEIKVNNLEHLEESAK 654
 QY 761 SEFKPMQNEQKEVYSKFRCTKINLLATTAEGEGLDKECNIVIRGLVYNEIAYVQAR 820
 DB 655 OKFGEITTYQKEIEDK-----KISEENL-LEE--VEKAKVADAEVYKQKE 698
 QY 821 GRARDESTYVVA-----HSGSGVIEHETVNDFREKKMYKAIHCVQ-----862
 DB 699 IDKROQHIAEVAALMEKHQYDKIIEERD-----SELGLYKKSQEQDSLSRAS ETELS 754
 QY 863 NKKPEEVA-HKILELOMOSIMEKKMKTKRNIAK-HYNNNSLITFL 906
 DB 755 NLKAEILLSVKKOLEIEREKELKLRKAENYATLKEKKDKKTQYFL 800

RESULT 10
 US-09-308-375-2
 : Sequence 2, Application US/09308375
 : Patent No. 6300117
 : GENERAL INFORMATION:
 : APPLICANT: Genencor International, Inc.
 : TITLE OF INVENTION: Proceases From Gram Positive Organisms
 : FILE REFERENCE: GC394-PCT
 : CURRENT APPLICATION NUMBER: US/09/308,375
 : CURRENT FILING DATE: 1999-05-14
 : EARLIER APPLICATION NUMBER: EP9719636.4
 : EARLIER FILING DATE: 1997-09-15
 : NUMBER OF SEQ ID NOS: 3
 : SOFTWARE: FastSeq for Windows Version 3.0
 : SEQ ID NO 2
 : LENGTH: 2285
 : TYPE: PRT
 : ORGANISM: Bacillus subtilis
 : US-09-308-375-2

Query Match	Similarity	3.2%	Score 171.5	DB 4	Length 2285
Best Local Similarity	18.2%	Pred. No. 3.5e-05			
Matches 214	Conservative 177	Mismatches 392	Indels 393	Gaps 50	
QY	15	ISCFARVAKMYQVPEVLDYLTFLFAEYKEQIQRTVATSGNNCAVELLSLEKGVNHDG	74		
DB	353	ISGLKEVNSQAEIDTLMTNI-----RVMNMPQVKYKELLQPSID----	LG 395		
QY	75	WT--REVEALRRGSPPLAARYMNPETDLPSPSENAHDEYLLNLNLITF-----TLVD	127		
DB	396	DLSMKRIDIDQMTDQFGRMGFDESELSITLTFTA-----QVLAIVSDILPDDIVNTLTA	449		
QY	128	KIL-----VRD-----VLDMKMEBE-----LTTEDNRIRIAAENNUNSGFHILKRIYQKE	175		
DB	450	AMLNFNFIANDSISTADKLNEVDNNAVTTLLDANSIRKAGSTAFEGV--LNDL----	504		
QY	176	NMFSAFLVNLROTGN-----	NEIQV 195		
DB	505	GTTTAAIATTRSGNIYVNSLKTIFARIQNNOSSIKALPQIGLSYKTAGTAKKSADLIS	564		
QY	196	ELTG-----SDCSSEN-----AEIENLSQVDPQ-----VEEQL 224			
DB	565	EVAAGKMDTLSDAKONTSIGVAGITQVLSRFNMMNNFSTAQNAKTAAN--GSAMGEQVK	624		
QY	225	LSTYQPNLEKRYVMENNSSSESSPADS--SVYSESTSLAEQSVSLDENJHNSMGSD	283		
DB	625	YADSIQARVNK-----LONNTEFAIQAASDAFISDGLIEFTQAASDLNA--GVYKSVGRL	680		
QY	284	SCFMGSDSD-----DENVARASPEPELDRLRYQMEVAPALHKKILITPTGSKIV	348		
DB	661	PPLAAVSTATLLEKKNRTTLAS-----SLILGTBAW-----QDTLTAISLEAGNTA	729		
QY	339	AVYIAKHLDKKKKASPEKQYIVLVNKKYVLBOLEFKEPQPLKWKRYHGLSGDQLEK	398		
DB	730	AV-----ASRVLKTALRGILVSLVGAF-----AAVGALESLS-- 765			
QY	399	SPEPVAKSCDIIITIAQILLENGLNENGEDAGVOLSDSLIIDECRIKKEAVYNNIM	458		
DB	766	SFAEKKKKKND-----DFEGSOQTVN-----EATIKNDSTKRLQ 800			
QY	459	RHYLQKLNKNR--LKKENKPVLPDQIIGLASPCVGATKAKAEEHJIKLANLOAF	516		
DB	801	QIKELQKKVKSLSLSDDEQETLOYTQOLAQTFFPALVKYKIDSCJAN-----LTKNKLLEK--	856		
QY	517	TKTYKENIDQLKNOIQEPPKFFALADATREDPEFKLEIMKQTYQNSPMSPECTO	576		
DB	857	ALENNKEYIALKKQFRODAKK-----TFEDASKR--IKSKNHLKQYK--ADYNKGR	909		
QY	577	PYEQWAIQMEKKAARKGNKREYCAE-----HLKRYNALQINDTIRNIAVTHLEFEY	630		
DB	910	KMDLADDDDDYKVAADKAKQSMILKQSDIESGNAKYKDSVLSIANNYSSHLSNTLKTISI	969		
QY	631	NEKKKKKRVIEDDD-----HGGHLE-----Y 653			
DB	970	SOVAVNK--LNLKDDLDLPELEKFFSSSLQLOEKMKQALNSGGKAKAFDNA<4>DLSILEEY	1027		
QY	654	CGDEDEDLKKPL-----KLDETDLEPLMTLFFENKKMKLRLA--NP--ENEKELIFK--	703		
DB	1028	SKSDSDIIVFKKSPDKAQKNINDGDKSLSSVASEVDDLEETLALAGNEA--LGGKKIKREAL	1087		
QY	704	-----LKNITME-----QYRTRESAGIIFTKTROSAVA--DMITENIKFAEYGVVA	791		
DB	1088	DANSYDDIKAAIKESMDAMQFSDQVNLGDIFFNNIKDQVAPINDLL--LKMA-----	1138		
QY	752	HHLLIAGHSEKPKPMQNOKEVYSK-----PRTKINILATITV<4>EGLDIKEEN	802		
DB	1139	-----EKXSISANEANTLLQDKELAQASISLENCVYKINRDEVIAKKVKKLDAVN	1188		
QY	803	IVIRYG--LVTRKEI-----AMVOARGRARADESTYVIAV--SSGVIIEET	845		
DB	1189	DMVTSNKILMKREVNNAIKTILADTLRISLKLKREKERLDMKSEAL-----SDLEFVS	1242		
QY	846	VND-----PREKNMY-----KAIHGVONMK-----PEEVAKILELOMOSSI	891		

```

Db      1243   INNADAKKELKKLEFKMKLPGGYSNSOIEAMQSVKSLSEYISASBEALSTCHMKAOL 1302
QY      882 ME-----KKMKTRKNIAKHKKNNPSLTFLCKNCVSLACSGLELHWIEK 925
Db      1303 VEAGTSLFNMTDQQEKANEETKTSMYVDVKY-----EALFK 1339
QY      926 MH-----HYNMTPPEKELY---LYENKALOKE 950
Db      1340 VNAELDKYNKVNDPKYSOKRYDAIKKAIKALOOK 1375

RESULT 11
US-09-134-001C-5080
: Sequence 5080, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
: TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GTC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ ID NOS: 5674
: SEQ ID NO 5080
: LENGTH: 3696
: TYPE: PAT
: ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5080

Query Match          3.1% Score 167; DB 4; Length 3696;
Best Local Similarity 19.5%; Pred. No. 0.00018;
Matches 200; Conservative 153; Mismatches 394; Indels 278; Gaps 44;

QY      41 EVKEOIORTVATSGN-----MQAVELLSTLEKGVML--GMTREFEVALRRTGSPLAA 92
Db      2641 EKANOQOSTIATHPSTIEBREQASAKLOEVILKKAIAKIDKGOTNDVEKTVANNAIEIE 2700
QY      93 RYMNELTDLPSPSEMAHDEY--LQLNLLOPIYLVDKLIVROVLCKMEBELLIIE--- 147
Db      2701 NLTPATTYDKAKAADVANEKEQKNLQINSNDEATEEFKLVASDNINHVETTNCALIEDAP 2760
QY      148 DNRRIAAFMNNGESGVRLIKRIYOK-----ENMSAFNLNVLRQGN----- 190
Db      2761 DTNOVVNEVENKIGT-IKDIOPLVVRKKPPANSKIESAVEKKTEITNQTNATHLEVREG 2819
QY      191 NEL--VOELTGSDC--SESNAEIEMLSOVDPQVEEQGLSTTVOPNLEKE-----WVG 239
Db      2820 NOLNJOHEAKMDVNOSQTNOQYENAEQNSLDQI-----KNFRDFSKKANVALEYIKA 2873
QY      240 MENNSSSESFADSSVVSSEDTSILAEGSVSCIDESLGHSNMGSDGTGWSDSDEAVAAK 299
Db      2874 QONKIDEIQEFSATQEEDNALQH-----LDEQYKETIINSIGNANT--DNEVINAKT 2924
QY      300 ASPPELQURPYQMEVAQPALEGKNIITCLPTGSGKTRVAYIADHDLDKKKKKSPEOKV 359
Db      2925 -----SGLNNTIEY--RPEVNRKRNA----- 2943
QY      360 IYLVNKVLLVEQLFKEKQEPFLKKWYRVIGLSGTQDLKI--SFEVYVKSCLII--IAQIL 417
Db      2944 -----TLKLYDY--SDTOEALINYIPATED-----ELQ 2970
QY      418 E-NSLIN--LENGEDAGVQLSDFSLLIIDECHTHNKA--VYNNIMRYLMOK KNNRL 471
Db      2971 EANSKLNKILTDAKKQIGLA-----HTNNEVDVIYNEV---SOKWKT--- 3009
QY      472 KKKNRPVIRPOLIGLITASPGVGATKOAKAEFHILKLCANDLAT-----IKTYEN 524
Db      3010 -----ILPRVDTKAARSAVNLNALAQ-LIKTFENTADVTHERENDAINHVEQ 3056

```

```

QY 525 LQDLMQIOEPCKKFAIA-----DATERDPFKKILLEIMFIQTGCONSP 569
QY 3057 LSLVFNATIEKDRKDIOVADDELFGLELNSIFINITQKPTARKA'SGMASQUNNSINNIP 3116
QY 570 MSDFGOPYROMAIOEMKKAAGKKNRERVCACHLKKYNEALUJNITIVIAATYLEIF 629
QY 3117 ---YATBEERQIALNKVKALVUDANEK-----IREANTDSFVAST--KINAITILLOAI 3164
QY 630 YNEKOKKRAVIEDSDDEGDDDEYCDGDEDEDLKKPLULEDTGRLMTI'FENNKMILKR 689
QY 3165 SAUVOKPAQAFELINQAOEIQERINGSNDATREKEKALQVH'LVNHSITINNVMK- 3223
QY 690 LAENPEYENK-----LTKLRNTIMEQYTRTESAK'IT-----ITKTROSANA 733
QY 3224 ---NOEYVDTKDKTIEAIHKIKPISIKPOLANEITIQIDTOPH'IKNNKISTVEEKASA 3280
QY 734 LQMITENEFKAEVGVKAAHHLIGAGHSSEFKPMTONEQKEVINSNHTGKINILIAATVAH 793
QY 3281 IDKLITARIATAIDKAO-----TNEEVKNIKKJSIDELSKILPYIEIK 3325
QY 794 EGI--DIKECNIVIRGLVT-NEIAMVQARGARADES'YVLVNI-SGSVIEHETVNDP 849
QY 3326 SAARNEIHOKEAVIR-GLINDNEEATKEEDIALNQLD'TLTQAVNSID'ATINEAVNKA 3384
QY 850 RE--KMMYKAIHGVQNMKKPEFYAHKILE-----QMSQIMKMKKK KNIKAKHK 897
QY 3385 KEIANSEINKISVIAIKRPPALA-ELIQELADKKLMPKOSQEA'FIEKO-SATINELQOLK 3443
QY 898 NNSLITFLCKNSVLACSGEDIHVEKMHVNMPEFELYIVRNKA'QKCK-----A 952
QY 3444 SAINNHIOSONNESVSALAKESISLIDIE-----LAKHKKIPAKATYIDGS 3490
QY 953 DYOIN 957
QY 3491 DOKIN 3495

```

```

RESULT 12
US-08-938-105-3
Sequence 3, Application US/08938105
Patent No. 6353151
GENERAL INFORMATION:
APPLICANT: Ielwand, Leslie A.
APPLICANT: Vikstrom, Karen L.
TITLE OF INVENTION: TRANSGENIC MODEL FOR HEART FAILURE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross P.C.
STREET: 1700 Lincoln St., Suite 3500
CITY: Denver
STATE: CO
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.40
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/938,105
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Crook, Wanneil M.
REGISTRATION NUMBER: 31,071
REFERENCE/DOCKET NUMBER: 3595-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1886 amino acids

```

```

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-938-105-3

```

```

Query Match 318; Score 166.5; DB 4; Length 1886;
Best Local Similarity 18.4%; Pred. No. 6.9e-05;
Matches 184; Conservative 167; Mismatches 367; Indels 241; Gaps 37;

```

```

QY 114 YLQNLNLOPTLVADKL-----VRDVLQK-----MEELLT-IEDRN-----RIA 153
QY 782 YKIKPILKASATEKEMANMKKEFGVADALEKSAARKKELEKKNVSLQEKNI'LOLOVQ 841
QY 154 AAEENNGNSGVN--ELKRIYOKENMFSAPLVNL'PQGNNEIYQDELTC-----IXSES 205
QY 842 ARODMLDAEEKCDQDILINKKIQLEAKVEMTE--FLPEFEFENNAITLAKKKRI'F'ECSEL 899
QY 206 MEIENLSQVDSPOVEQDOLSTTVQPNLEKVEKMGENNSSSESPADSSVSESISLAE 265
QY 900 KKDIDDELTLAKVEKHNATEKNKYNLTIEKAGLDEIIAK-----LTKK'ALOEA 951
QY 266 SVSCLEDSIGHNSNMGSDSGTWGSDDEENVAARASPEPELQRPYQMEVAPQ'EGKN 325
QY 952 HQQALDD-----LQAEKRVNLTITKSKVYLEQVODLEGSLEQ----- 989
QY 326 IICLPFGSGKTRVAVYIAKDILDKKKAASPEQKVIVLVNKLVLQ'LFKPEFQ'IKKMY 385
QY 990 -----EKVYMDLERAKRKLEGLKLTQESIMDLENDKQLEKELKKKEF----- 1034
QY 386 RYIGLSGDTQIKISPEVYKSGDIIISTAOITENSLNLE-----NGSDACVQI'SDPSLI 440
QY 1035 -----DISQONSKIEDEQALALQOLKUKENQAKITELIEEL- 1071
QY 441 IDECHHTNKEAVYNNIMRHYLMQIK--NNRLKKNKPVIPDQIIGLTASPCV'GATK 498
QY 1072 ---EAEPTARAKVER--LRSDLTRELEISELEF-----AGAGTS 1107
QY 499 -----QAKAEHITLKCANLDAFTIK-----TYKE-----NLDLQKNOI 533
QY 1108 VOIENMKKREAEFOKMRDLEBATEALQHEATAAALRKKHADSVAEIGEOIDIM'OKYKOLE 1167
QY 534 EPCKKPAI-----ADATREDPKKLEIMTPIOTYCQSPMSD 572
QY 1168 KKSSEKLELDVDTYSHMEOI'IKAKANLEKVSRTLEDQNE-----YVKLEEA'KSLND 1221
QY 573 FGTQPYE-----QWAIOMEKKA-----KKGNRERVCACHLKKYNEALQIN'IRMD 621
QY 1222 FTTQRAKIQTEGELARQLEEKALIMQUTRGKLSYTOQMEDLKR-----QLEF'WKKKN 1276
QY 622 ATTH-----LETFYNEKDKKRAVIEDSDDEGDDDEYCDGDEDDPDK'KKPLKD 670
QY 1277 ALAHALQSAHDCDLLREOYEEMKAKELQKRVLSKANSEVAQWTKYETQAL'KTEELE 1336
QY 671 ETDRIPLMTL'FENNKMILKRL--AENPEYENKILT-----KLNTIMEQYTRTES-- 718
QY 1337 EAKK-----KLAORLQDAEAVAVANAKKCSLEKTKHRLQNE'LEDLAMI'VHRSNA 1386
QY 719 ANGILFTKTROSAYAL'SOMITENEFKAE-----GVGAAHHLIGAGH 759
QY 1387 AAAALDKQORNDKILAEH---KQKYEESQSELESQKEARSISTELFKLKNAYHESLEH 1443
QY 760 SSEFPMTQNEQKEY-----ISKR-----TGKINILIAATVAH'KJIDIK 799
QY 1444 LETFKRENKNIQEEISDLTEQLEGSGKNVHELEKIRKOLEVEKLELQSALEEAFV'SLEHE 1503
QY 800 EBNIVIRGLVNIENIAMVQARGARADESTYVLVHSSQYIEHETVNDP'FEKKMYKAIH 859
QY 1504 ECKI--LRQOLEFQIKAEIERKIAKDEMPQAKRNHILRVYDSQTSIDAE'FRKNEELR 1552
QY 860 CVQNMKPEFYAHKILELQMSIMERKMTKRNIAKHYNNNSLITFLCKNSCV'V'SG-- 917
QY 1563 VKKKHEGDLNEKEIOLSOANRIASLQAKHLKNAQAHKLD-----TOLQDDAVRANDDLK 1617

```


Best Local Similarity 19.2%; Pred. No. 0.00014;
Matches 215; Conservative 178; Mismatches 423; Indels 304; Gaps 48;

QY	22	VKNITQVEPVLDTLTPLEAVKQIORTVAATSONNAVILLISLITKEGVWUHGRTREVE	81
Db	832	MKLYTFKIKPLIK-----SATEKEEMATMEPGRIT-----PFI EKS-----E	869
QY	82	ALRTGSPLAARTINPELIDLPSPSTENADELYOL-----LNL LO---PILVOKLVR	132
Db	870	ARRK-----ELEKRVSLLOEKNDLOLOVQAHODINIABEEKPOLLINKIOL	917
QY	133	DVLDCMEELLTIEDENRITAAENNGNESGVELLKRT-----VLAKNMEASFLN	183
Db	918	EAYKKNNEELDEEENENALTKAKKKRLEDECSSELKKDIDDLPI PLAKTRFK--HATHN	975
QY	184	VLR-----OTGNNELOVELTSGDCSSNAEINLSO-VICPOVFOLLSTVQV--NLEK	235
Db	976	KVKNLITEEMAGLEI IAKLT-----KEKKALOEHQCALIDLOVTEIDKVAISKSVKJLEO	1031
QY	236	EVMGNENSSSPADSSVYS-----ESPTSLAEGSVCLIDENI JHNSN-ANDSTCMISD	290
Db	1032	QVDDLOBSLODEKVRKNDLEBARKKIEGLDKLTQESIMLE-----NDK VLEELKKE	1088
QY	291	SD-----EENVAHASPPELOLIRPYQMEVAQPALHX-----K	323
Db	1087	FDIQONSKIEDEQALALQLOKKLKENOARIEBELBELAEERAKVAKYKLSDSRELE	1148
QY	324	NIILCLPTSGKTRVAVYIAKOH-----LD-----KAKKASERGVI	360
Db	1147	EISRELEAGGATSVQIENKKKKRAEPQKKRRDLFEATLOHEVIAALKKKHAUSVAELG	1206
QY	361	VLVAKVVLVEQDLFRKEFOPLKRWYVIGLSGTQOLKISFPEVYKSCDIITSTAOILNS	420
Db	1207	FOIDNLOVRKQLEKE-----KSEKRLDUDVITSMEDIKAKANLKV	1250
QY	421	LNLNENG-----EPAQVQLSDFS-----LIIIECHRTNENAVYNNIR	459
Db	1251	SRTELEADANEVRYKLEBAORSINDPTTORAKIOTENGELAROFER-----REALISOLTR	1305
QY	460	HYL-----MOKLNKNRLEKKNKRVIPULPOLIGTASPVGATK-----	496
Db	1306	GKLSYTOQMMDIK-ROLEEBGK-----ANNALAHMLJASRHPDILBROYDEE	1355
QY	499	-QAKAEHILKCANLDATFIKTIVKE-NLDOLKNOIOENCKFR--IADP REDEFKKL	554
Db	1353	TEAAAELORYLKSANSEVQAOMRTKYETDAIORTBELEEKKKLVARLOQAD-----	1403
QY	555	LEIMTRIOYCOMSPMDFCTO-PYQOMAIOMKKKA-----KGNKRTFWCAEHLKY	607
Db	1404	-EAVEAVNACSSLEKTKHLOMEIEDIANDVERSAALMALIKKORNNIKLAKMKKY	1462
QY	608	NEALQ-----INDTIRIDAY-----THLETFNEBKOKKFAVINDSDGND	651
Db	1463	EESOSELESSOKEARSLSTELFKLMAYESLEHLETFKENNINQ-----FISLO-TE	1516
QY	652	EYCGDDEDDDKKPLKIDTDRFLM-----LPIENKKMLKRLATNREVENKLT	702
Db	1517	QJEGGKNGHELEVRKQLEVERLEQSALEBEAEASLEHEBECKILRAOELIJOIAETER	1578
QY	703	KL-RNTIMEQYRTTESAGIIFITRTROSAVALSOMITENEKFAEYGVKMHILGACHS	760
Db	1577	KLAKDEDEMOQRNHRQVAVDSLOTSIDAETSRNEVLKVKKNH--GDN NEMETOUSHA	1634
QY	761	SEFEPHTQNOQKEVYISKFRIGKINLLIATTVABEGIDKECNI VI---RGLAVNEI---	814
Db	1635	NRMMAEPKOKVYKSLQSLTKDQTOIDDAAVRAND--DLKE-MIVATERNNMLQLEBEL	1690
QY	815	-AMVQAGRRAR-----ADESTVVLVAHSQGVTEHETVNDPREKKMKALIHQVONM	864
Db	1691	RAVEYQERBRKLADEOELLITTSRGVQLHSQNTSLINOKKKH--DOLTQUL-NEVEAVAYCB	1750
QY	865	K-PEYVAHKL--ELOMOSIMEKKMKTKRNIAXHYKNNPSLITFLCKNDSVLAGSGEIH	921

```

Db 1751 RNAEEKKKKKILTDAAAMAAEELKEEDOTSALHEEMKKMMQOTIKDL-----1795
QY 922 VIEKMHVNTPEFKELIYRENKALQK-KCADYQINGEL 960
      | : | : : | | | : : : | :
Db 1796 ----QHRLD---EAEQALALGSKKKQIQKLEAVRVELEGEL 1828

```

RESULT 15
US-09-134-001C-4452
; Sequence 4452, Application US/09134001C
; Patent No. 6380370

```

1  APPLICANT: Lynn Doucellette-Stamm et al
2  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
3  TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
4  FILE REFERENCE: GTC-007
5  CURRENT APPLICATION NUMBER: US/09/134,001C
6  CURRENT FILING DATE: 1998-08-13
7  PRIOR APPLICATION NUMBER: US 60/064,964
8  PRIOR FILING DATE: 1997-11-08
9  PRIOR APPLICATION NUMBER: US 60/055,779
10 PRIOR FILING DATE: 1997-08-14
11 NUMBER OF SEQ ID NOS: 5674
12 SEQ ID NO 4452

```

TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4452

[illegible]

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 15:44:58 ; Search time 81 Seconds

(without alignments)
1686.196 Million cell updates/sec

Title: US-09-515-363C-2

Perfect score: 5311
Sequence: 1 MSGXSTDENRYLISCRA.....LPIFFPNLDSYGLFSDSD 1025

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002:*

```
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5311	100.0	1025	AAE10155	Human melanoma dif
2	5285	99.5	1025	AAE10155	Human RNA helicase
3	4306	81.1	838	AAE10165	RNA helicase conse
4	2595.5	48.9	558	AAU23090	Novel human enzyme
5	1860	35.0	417	AAU23647	Novel human enzyme
6	1839	34.6	356	AAU40129	Human polypeptide
7	1769	33.3	348	AAU41915	Human polypeptide
8	1344	25.3	678	AAU93708	Human protein sequ
9	1093	20.6	304	AAU00296	Interferon induced
10	1093	20.6	308	AAU00298	Interferon induced

11	861	16.2	267	22	AAU00297	Interferon induced
12	861	16.2	271	22	AAU00299	Interferon induced
13	808	15.2	448	22	AAU23654	Novel human enzyme
14	783	14.7	166	22	ABR42219	Peptide #9725 enco
15	783	14.7	166	22	AAU63104	Human brain expres
16	783	14.7	166	22	AAU75915	Human bone marrow
17	783	14.7	166	22	AAU64026	Peptide #10063 enc
18	782	14.7	154	22	AAU64314	Human colon cancer
19	782	14.7	154	23	ABP41112	Human ovarian anti
20	666.5	12.5	357	21	ABR42961	Human OMF ORF2745
21	563.5	10.6	447	23	ABP41171	Human ovarian anti
22	558.5	10.5	447	22	AAU65485	Human protein sequ
23	521	9.8	312	22	AAU23099	Novel human enzyme
24	488	9.2	92	22	ABR27911	Human peptide #562
25	488	9.2	92	22	ABR33083	Peptide #589 enco
26	488	9.2	92	22	ABR18531	Protein #550 enco
27	488	9.2	92	22	AAU63880	Human brain expres
28	488	9.2	92	22	AAU6268	Human bone marrow
29	488	9.2	92	22	AAU4138	Peptide #572 enco
30	488	9.2	92	22	AAU6544	Peptide #581 enco
31	488	9.2	92	22	AAU01876	Peptide #558 enco
32	488	9.2	92	23	ABG35915	Human peptide enco
33	453	8.5	752	22	AAU6107	Putative P. abyssi
34	406	7.6	1909	22	AAU31166	Amino acid sequenc
35	395.5	7.4	764	22	AAU62028	Recombinant P. fur
36	357	6.7	71	22	ABR42510	Peptide #10016 enc
37	357	6.7	71	22	AAU76213	Human bone marrow
38	357	6.7	71	22	AAU63323	Peptide #10360 enc
39	332.5	6.3	1383	22	ABR63586	Drosophila melanog
40	332	6.3	1924	22	AAE09768	Human diacyl protei
41	324.5	6.1	190	22	AAU63792	Human immune/hema
42	298	5.6	55	22	AAU18482	Peptide #4916 enco
43	298	5.6	55	23	ABG40321	Human peptide enco
44	286	5.4	503	21	AAU90944	Cenarchaeum symbio
45	284	5.3	502	21	AAU90928	Cenarchaeum symbio

ALIGNMENTS

RESULT 1	AAE10155	standard; Protein: 1025 AA.
ID	AAE10155	
AC	AAE10155;	
XX		
DT	29-NOV-2001 (first entry)	
XX		
DE	Human melanoma differentiation associated (Mda)-5 protein.	
XX		
KW	Human; melanoma differentiation associated gene; Mda-5; interferon; IFN;	
KW	RNA helicase motif; caspase recruitment domain; CARD; therapy; melanoma;	
KW	neuroblastoma; astrocytoma; glioblastoma; multifactor; cervical;	
KW	breast; colon; prostate; osteosarcoma; chondrosarcoma; systemic toxicity;	
XX	central nervous system; cytosolic; apoptosis..	
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	Region	125..174
FT		/note- "Reveals sequence homology with other CARD
FT		proteins"
FT	Region	722..823
FT		/note- "Reveals significant homology to RNA helicase
FT		C-terminal conserved domain"
FT	Domain	331..336
FT		/note- "ATPase A motif"
FT	Domain	443..446
FT		/note- "ATPase B motif"
PN	WO200164707-A1.	
XX		
PD	07-SEP-2001.	

XX 28-FEB-2001: 2001WO-US06960.
 XX
 XX 29-FEB-2000: 2000US-051363.
 XX
 XX (UYCO) UNIV COLUMBIA NEW YORK.
 XX
 XX Fisher PB, Kang D, GopalKrishnan RV;
 XX WPI: 2001-565494/63.
 DR N-PSDB: AAD17203.
 XX
 PT Nucleic acid sequences encoding a Melanoma Differentiation Associated
 PT Gene useful for cancer cell growth suppression, apoptosis and
 PT anti-viral activity
 XX
 XX Claim 21: Page 18-19; 152pp: English.
 PS
 CC The present invention relates to an isolated nucleic acid encoding a
 CC melanoma differentiation associated gene-5 (mda-5) polypeptide. Mda-5
 CC contains a caspase recruitment domain (CARD) and a RNA helicase motif.
 CC Mda-5 is a novel interferon (IFN) inducible gene with structural
 CC similarities to RNA helicases and CARD motif containing proteins. Mda-5
 CC is induced during terminal differentiation in human melanoma cells
 CC treated with the combination of recombinant fibroblast IFN and the
 CC anti-leukemic compound mezerein (MEZ). Mda-5 is useful for identifying
 CC compounds that may induce its expression. Mda-5 is useful for treating
 CC cancer e.g., melanoma, neuroblastoma, astrocytoma, glioblastoma
 CC multifactor, cervical cancer, breast cancer, colon cancer, prostate
 CC cancer, osteosarcoma, chondrosarcoma and especially where the cancer is
 CC a cancer of the central nervous system and apoptosis. The mda-5 promoter
 CC exhibits melanocyte tissue specificity and minimises systemic toxicity.
 CC The present sequence is human mda-5 protein.
 XX
 XX Sequence 1025 AA:
 SQ
 Query Match 100.0%; Score 5311; DB 22; Length 1025;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1025; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 481 LPQILGLTASPGVGATQAKAEERILKLCANLDAFTIKTVENILQDKNOIQEPCKKFA 540
 OY 541 IADATREDPFKEKLEIMTRIQTQYQMSPMDSFGTQPEYQMAIQMEKKAARKGNKEKRV 600
 DB 541 IADATREDPFKEKLEIMTRIQTQYQMSPMDSFGTQPEYQMAIQMEKKAARKGNKEKRV 600
 OY 601 AEHLKRYNEALQINDTIRMIDAYTHLETFYNEEKDKRAVIDEDSDGDEYDCQDEDE 660
 DB 601 AEHLKRYNEALQINDTIRMIDAYTHLETFYNEEKDKRAVIDEDSDGDEYDCQDEDE 660
 OY 661 DDLKRPKLDETDNPLMTLFFENNMLRLAENPEYENKLTJLNTIMEQYTRRESAR 720
 DB 661 DDLKRPKLDETDNPLMTLFFENNMLRLAENPEYENKLTJLNTIMEQYTRRESAR 720
 OY 721 GIIFTRQSAAVALSOWITENKEFAEYGVKAAHHLIGACHSSEFKPMONEKEVYSKRT 780
 DB 721 GIIFTRQSAAVALSOWITENKEFAEYGVKAAHHLIGACHSSEFKPMONEKEVYSKRT 780
 OY 781 GKINLLIATVAEEGLDKECHIVIRGLVNEIAMVQARGARADESTYVLVAHSGSV 840
 DB 781 GKINLLIATVAEEGLDKECHIVIRGLVNEIAMVQARGARADESTYVLVAHSGSV 840
 OY 841 IEHETVNDPREKMYKAIHCYQNMKPEYAKHLELOQSIIEKKMKTKRIAKHYKNP 900
 DB 841 IEHETVNDPREKMYKAIHCYQNMKPEYAKHLELOQSIIEKKMKTKRIAKHYKNP 900
 OY 901 SLITFLCKNCSYLACSGEDIHVIEKMHVNMTPFEKELYIRENKALOKKADYQINCE 960
 DB 901 SLITFLCKNCSYLACSGEDIHVIEKMHVNMTPFEKELYIRENKALOKKADYQINCE 960
 OY 961 ICKCGQAGCTMMVHHGDLPLCKIRNFYVFKNNSTKKQYKWEPLTTPFNLCYSECT 1020
 DB 961 ICKCGQAGCTMMVHHGDLPLCKIRNFYVFKNNSTKKQYKWEPLTTPFNLCYSECT 1020
 OY 1021 FSDPD 1025
 DB 1021 FSDPD 1025

RESULT 2
 ID AAM47798 standard; protein: 1025 AA.
 XX
 XX AAM47798:
 DT 05-MAR-2002 (first entry)
 XX
 DE Human RNA helicase RH116.
 XX
 KW Human: RH116; RNA helicase; cytosolic; vifucide; anti-HIV;
 KW immunosuppressive; immunostimulatory; antitumoric; antiarthritic;
 KW antiarteriosclerotic; osteopathic; antidiabetic; hepatotropic;
 KW antiinflammatory; cancer; infection; HIV; hepatitis; genetic disease;
 KW autoimmune disease; graft rejection; vaccine.
 XX
 OS Homo sapiens.
 XX
 PN W0200185955-A1.
 XX
 PD 15-NOV-2001.
 XX
 XX 11-MAY-2001; 2001WO-FR01441.
 XX
 XX 11-MAY-2000; 2000FR-0006030.
 XX
 PA (ISTA-) ISTAC.
 PA (INSP) INST PASTEUR LILLE.
 PI Bahr G, Cocude C, Capron A;
 XX WPI: 2002-082898/11.
 DR N-PSDB: ABA04908.


```

XX New polypeptide, useful for treating and diagnosing cancer or
PT inflammation, and drug screening, comprises a human polynucleotide
PT homologous to RNA helicase
XX
XX Claim 1; page 89-93; 114pp; French.
XX
XX The present sequence is the protein sequence for human Rhl16. Rhl16 is a
CC 116kDa protein and has homology to RNA helicases (DExH box). Rhl16 and
CC its coding sequence are useful for treating cancer; active or chronic
CC infections (especially by HIV or hepatitis B or C); inherited genetic
CC diseases; (auto)immune diseases (particularly rheumatism, arthritis,
CC arteriosclerosis, osteoporosis and diabetes, but many others listed) and
CC to prevent graft rejection. Rhl16 and its coding sequence are also useful
CC for inducing, or increasing, the immune response to a vaccine.
XX
XX Sequence 1025 AA:
SQ
Query Match 99.5%; Score 5285; Dh 23; Length 1025;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1020: Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 1 MSNGYSTDENFRYLISCFRARKYMYIQVEPVDLYTLFLPAEYKQIQRTVATSGNMQAVE 60
DB 1 MSNGYSTDENFRYLISCFRARKYMYIQVEPVDLYTLFLPAEYKQIQRTVATSGNMQAVE 60
QY 61 LLISTLEKGVNHLGWTREFEALRRTGSPLAARYMPELIDLPSPFENAHDEYLOLLNL 120
DB 61 LLISTLEKGVNHLGWTREFEALRRTGSPLAARYMPELIDLPSPFENAHDEYLOLLNL 120
QY 121 LQPTLVKLLVROVDLCKBEELLITIEDNRRLAAENNGESGRLLEK:VQENMFSA 180
DB 121 LQPTLVKLLVROVDLCKBEELLITIEDNRRLAAENNGESGRLLEK:VQENMFSA 180
QY 181 FLVNLQGTGNELVQELTSDCSSESNABTEBNISQVDPQVEBOLLSTLYPLNEKVCWM 240
DB 181 FLVNLQGTGNELVQELTSDCSSESNABTEBNISQVDPQVEBOLLSTLYPLNEKVCWM 240
QY 241 ENNSSSSFADSSVSESPITSLAEGSVCLDESLGNSNMWGSNGSGMSQDEENVAARA 300
DB 241 ENNSSSSFADSSVSESPITSLAEGSVCLDESLGNSNMWGSNGSGMSQDEENVAARA 300
QY 301 SPEPELQLRPYQMEVNAQPALEGNIIITCLPTSGKTRVAVYIAKDHDKKAKSEPGKYI 360
DB 301 SPEPELQLRPYQMEVNAQPALEGNIIITCLPTSGKTRVAVYIAKDHDKKAKSEPGKYI 360
QY 361 VLVNKKVLLVEQLFRKEPQPLKRWYVIGISGDTQLKISPEVYKSCDII:STAQILENS 420
DB 361 VLVNKKVLLVEQLFRKEPQPLKRWYVIGISGDTQLKISPEVYKSCDII:STAQILENS 420
QY 421 LNLNENGEGAGVQLSDPSLIIIDECHTNKAEAYNNIMHLYLMOKLKNK:IKENKPVIR 480
DB 421 LNLNENGEGAGVQLSDPSLIIIDECHTNKAEAYNNIMHLYLMOKLKNK:IKENKPVIR 480
QY 481 LPQILCLTASPGVGAATKQAKAEHLIKLCANDAFITIKVENLQOLKNQ:IDEPKPKRA 540
DB 481 LPQILCLTASPGVGAATKQAKAEHLIKLCANDAFITIKVENLQOLKNQ:IDEPKPKRA 540
QY 541 IADATREDPEKLEIMTRIQTYCOMSPSDFGTOPYLMAIOMEKKAARERKRYC 600
DB 541 IADATREDPEKLEIMTRIQTYCOMSPSDFGTOPYLMAIOMEKKAARERKRYC 600
QY 601 AEHLRKYNALQINDTRIMDATHTLETPENEKOKKFAVIDEDSDGG:4:EGCDDDED 660
DB 601 AEHLRKYNALQINDTRIMDATHTLETPENEKOKKFAVIDEDSDGG:4:EGCDDDED 660
QY 661 DDKKPLKDETRFLMTLTFEENNMKRLAEPEYENELTKRLMTIM:GYRTRESAR 720
DB 661 DDKKPLKDETRFLMTLTFEENNMKRLAEPEYENELTKRLMTIM:GYRTRESAR 720
QY 721 GIIFTKRQSAVALSOMITENEKFAEYGVKAHHLIGAGINSSEKPMTONH:KEVISKPP 780
DB 721 GIIFTKRQSAVALSOMITENEKFAEYGVKAHHLIGAGINSSEKPMTONH:KEVISKPP 780

```

```

QY 781 GKINLLIATVAEEGLDIKECNIVIRGLVTNELMAYQARARADESTYVA:HSNSGV 840
DB 781 GKINLLIATVAEEGLDIKECNIVIRGLVTNELMAYQARARADESTYVA:HSNSGV 840
QY 841 IERTVNDFEREKMYKAIHCVQNMKPEEYAHKILELOSMIMEKMKTKRNIAK:YKNNP 900
DB 841 IERTVNDFEREKMYKAIHCVQNMKPEEYAHKILELOSMIMEKMKTKRNIAK:YKNNP 900
QY 901 SLITFLCKNSVSLACSGEDIHVIEKMHVNTPEPEKELIYREKALOKCAV:YINCEI 960
DB 901 SLITFLCKNSVSLACSGEDIHVIEKMHVNTPEPEKELIYREKALOKCAV:YINCEI 960
QY 961 ICKCGAMGTMMVHKGDLPCIKINPFVYFNKNSSTKKQYKMYELPTFPNL:DYSECL 1020
DB 961 ICKCGAMGTMMVHKGDLPCIKINPFVYFNKNSSTKKQYKMYELPTFPNL:DYSECL 1020
QY 1021 FSDSD 1025
DB 1021 FSDSD 1025

RESULT 3
AAE10165
ID AAE10165 standard; Protein; 838 AA.
XX
XX AAE10165;
AC
XX 29-NOV-2001 (first entry)
DT
XX
XX RNA helicase conserved motif of human Mda-5 protein.
DE
XX
XX Human: melanoma differentiation associated gene; Mda-5; interferon: IFN;
XX RNA helicase motif; caspase recruitment domain; CARD; therapy: melanoma;
XX neuroblastoma; astrocytoma; glioblastoma; multifome; cancer: cervical;
XX breast; colon; prostate; osteosarcoma; chondrosarcoma; systemic toxicity;
XX central nervous system; cyostatic; apoptosis.
OS
XX Homo sapiens.
XX
XX W0200164707-A1.
PN
XX
XX 07-SEP-2001.
PD
XX
XX 28-FEB-2001; 2001WO-US06960.
PF
XX
XX 29-FEB-2000; 2000US-0515363.
PR
XX (UYCO ) UNIV COLUMBIA NEW YORK.
PA
XX
XX Fisher PB, Kang D, GopalKrishnan RV;
PI
XX WPI: 2001-565494/63.
XX
XX Nucleic acid sequences encoding a Melanoma Differentiation Associated
PT Gene useful for cancer cell growth suppression, apoptosis and
PT anti-viral activity -
XX
XX Example 1; Fig 1D; 152pp; English.
PS
XX
XX The present invention relates to an isolated nucleic acid encoding a
XX melanoma differentiation associated gene-5 (Mda-5) polypeptide. Mda-5
XX contains a caspase recruitment domain (CARD) and a RNA helix use motif.
XX Mda-5 is a novel interferon (IFN) inducible gene with structural
XX similarities to RNA helicases and CARD motif containing proteins. Mda-5
XX is induced during terminal differentiation in human melanoma cells
XX treated with the combination of recombinant fibroblast IFN and the
XX anti-leukemic compound mezerein (MEZ). Mda-5 is useful for identifying
XX compounds that may induce its expression. Mda-5 is useful for treating
XX cancer e.g., melanoma, neuroblastoma, astrocytoma, glioblastoma
XX multifome, cervical cancer, breast cancer, colon cancer, prostate
XX cancer, osteosarcoma, chondrosarcoma and especially where the cancer is
XX a cancer of the central nervous system and apoptosis. The Mda 5 promoter

```

CC exhibits melanocyte tissue specificity and minimises systemic toxicity.
 CC The present sequence is RNA helicase conserved motif of human Mda-5
 CC protein.

XX Sequence 838 AA:

Query Match 81.1%; Score 4306; Dh 22; Length 838;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 LLLSTLEKGYMLGWTREFVEALRRTGSPPLAAYNPELIDLPSPSPFNHDEYLOLIM 120
 DB 1 LLLSTLEKGYMLGWTREFVEALRRTGSPPLAAYNPELIDLPSPSPFNHDEYLOLIM 61
 QY 121 LQPTLVKLLVRYDLDCKMEELLTEDNRRTAAAFNNKNSGVRELLKIVOKENWISA 180
 DB 61 LQPTLVKLLVRYDLDCKMEELLTEDNRRTAAAFNNKNSGVRELLKIVOKENWISA 120
 QY 181 FLAVLQGTGNNELVOELTSDSESNAEIENLSQVDPVPEBOLLSTTVLPILEKFWKIM 240
 DB 121 FLAVLQGTGNNELVOELTSDSESNAEIENLSQVDPVPEBOLLSTTVLPILEKFWKIM 180
 QY 241 ENNSSSSPADSSVSESOTSLAEGSVSCIDESLGHNSNMGSDGTMGSI SDEENVAARA 300
 DB 181 ENNSSSSPADSSVSESOTSLAEGSVSCIDESLGHNSNMGSDGTMGSI SDEENVAARA 240
 QY 301 SPPEQLRLPYOMEVAOPALEGNIIITCLPTGSGKTRAVAVYIAKDLHDKKKASEPGKYI 360
 DB 241 SPPEQLRLPYOMEVAOPALEGNIIITCLPTGSGKTRAVAVYIAKDLHDKKKASEPGKYI 300
 QY 361 VLVNKKLVYQQLFRKEFQPLKKWYVIGLSGDTQKISPEPVKSCDIIISAOQLIENS 420
 DB 301 VLVNKKLVYQQLFRKEFQPLKKWYVIGLSGDTQKISPEPVKSCDIIISAOQLIENS 360
 QY 421 LNLNLEGDAVQVLSDFSLIIDECHHTNKEAVYNNIMKHYYLMKKENKPVLP 480
 DB 361 LNLNLEGDAVQVLSDFSLIIDECHHTNKEAVYNNIMKHYYLMKKENKPVLP 420
 QY 481 LPQILGLTASPGVGATKQAKAEHLKLCANIDATFKIVKENLDOLKNOIOEPCKFA 540
 DB 421 LPQILGLTASPGVGATKQAKAEHLKLCANIDATFKIVKENLDOLKNOIOEPCKFA 480
 QY 541 IATATREDPPEKLEIMRIOTYGGMSPSDGTQPYMAIOMKKAAGKGRKREVC 600
 DB 481 IATATREDPPEKLEIMRIOTYGGMSPSDGTQPYMAIOMKKAAGKGRKREVC 540
 QY 601 AEHLRYNNEALQINDTIRMIDATYHLETFYNEEKDKKFAVIDEDSDGGIDEYCDGDEDE 660
 DB 541 AEHLRYNNEALQINDTIRMIDATYHLETFYNEEKDKKFAVIDEDSDGGIDEYCDGDEDE 600
 QY 661 DDLKKPLKIDETDFLMTLFEFENKMKLRLAEPYENKELTKLRNTIMCYTRTEESAR 720
 DB 601 DDLKKPLKIDETDFLMTLFEFENKMKLRLAEPYENKELTKLRNTIMCYTRTEESAR 660
 QY 721 GIIFTRGSAVALSOMITENEFKFAVGVKAHHLIGAGHSSEKPRPTONKQKEVISKPRPT 780
 DB 661 GIIFTRGSAVALSOMITENEFKFAVGVKAHHLIGAGHSSEKPRPTONKQKEVISKPRPT 720
 QY 781 GKINLLIATTVAAEGDIEKCNIVIRYGLVTNFIAMVQARGKARADESYVAVVAHSGSGV 840
 DB 721 GKINLLIATTVAAEGDIEKCNIVIRYGLVTNFIAMVQARGKARADESYVAVVAHSGSGV 780
 QY 841 IEHEVTNDRPREKMMYKAIFCYONMKPREYAHKLTLELOMOSIMKKKTKRNIAKATKN 898
 DB 781 IEHEVTNDRPREKMMYKAIFCYONMKPREYAHKLTLELOMOSIMKKKTKRNIAKATKN 838

RESULT 4

AAU23090 standard; Protein; 558 AA.

XX AAU23090;
 XX

DT 18-DEC-2001 (first entry)
 XX
 DE Novel human enzyme polypeptide #176.
 XX
 KW Human: oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
 KW lysase; hyperproliferative disorder; immunodeficiency disorder;
 KW autoimmune disorder; neurological disorder; metabolic disorder;
 KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
 KW blood-related disorder; infectious disorder; cytostatic; anti arthritis;
 KW nephrotropic; anticoagulant.
 OS Homo sapiens.
 PN W0200155301-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001W0-US01239.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0196125.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226270.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.

QY	654	CDGDEDEDLKKPKLPLETDEDTLFTFENNKKMLKRLAI:MPREYENELTKIRNTIMCYOT	713		
Db	361	CDGDEDEDLKKPKLPLETDEDTLFTFENNKKMLKRLAI:MPREYENELTKIRNTIMCYOT	420		
QY	714	RTEESARGIIPTKTRQSAVALSOMITENKRAEYGVAKMILIGAHSSEKRPMTONEOKI	773		
Db	421	RTEESARGIIPTKTRQSAVALSOMITENKRAEYGVAKMILIGAHSSEKRPMTONEOKI	480		
QY	774	VISFRGKIMLLATTVAEGDLKECNVIYIRGLVTNFLAI:VARGVARADESTYV	831		
Db	481	VISFRGKIMLLATTVAEGDLKECNVIYIRGLVTNFLAI:VARGVARADESTYV	537		
RESULT 5					
AAU23647	standard; Protein: 417 AA.				
XX	AAU23647;				
AC	18-DEC-2001	(first entry)			
DT					
XX	Novel human enzyme polypeptide #733.				
DE					
XX	Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;				
KW	lyase; hyperproliferative disorder; immunodeficiency disorder;				
KM	autoimmune disorder; neurological disorder; metabolic disorder;				
KW	inflammatory disorder; cardiovascular disorder; reproductive disorder;				
KM	blood-related disorder; infectious disorder; cytostatic; anti arthritic;				
KW	nephrotropic; anticoagulant.				
XX					
OS	Homo sapiens.				
XX					
PN	MO200155301-A2.				
XX					
PD	02-AUG-2001.				
XX					
PF	17-JAN-2001; 2001MO-US01239.				
XX					
PR	31-JAN-2000; 2000US-0179065.				
PR	04-FEB-2000; 2000US-0180628.				
PR	24-FEB-2000; 2000US-0184664.				
PR	02-MAR-2000; 2000US-0186350.				
PR	16-MAR-2000; 2000US-0189874.				
PR	17-MAR-2000; 2000US-0190076.				
PR	18-APR-2000; 2000US-0198123.				
PR	19-MAY-2000; 2000US-0200515.				
PR	07-JUN-2000; 2000US-0209467.				
PR	28-JUN-2000; 2000US-0214886.				
PR	30-JUN-2000; 2000US-0215135.				
PR	07-JUL-2000; 2000US-0216647.				
PR	07-JUL-2000; 2000US-0216880.				
PR	11-JUL-2000; 2000US-0217487.				
PR	11-JUL-2000; 2000US-0217486.				
PR	14-JUL-2000; 2000US-0218290.				
PR	26-JUL-2000; 2000US-0220963.				
PR	26-JUL-2000; 2000US-0220964.				
PR	14-AUG-2000; 2000US-0224518.				
PR	14-AUG-2000; 2000US-0224519.				
PR	14-AUG-2000; 2000US-0225213.				
PR	14-AUG-2000; 2000US-0225214.				
PR	14-AUG-2000; 2000US-0225266.				
PR	14-AUG-2000; 2000US-0225267.				
PR	14-AUG-2000; 2000US-0225268.				
PR	14-AUG-2000; 2000US-0225270.				
PR	14-AUG-2000; 2000US-0225447.				
PR	14-AUG-2000; 2000US-0225757.				
PR	14-AUG-2000; 2000US-0225758.				
PR	14-AUG-2000; 2000US-0225759.				
PR	18-AUG-2000; 2000US-0226279.				
PR	22-AUG-2000; 2000US-0226681.				
PR	22-AUG-2000; 2000US-0226868.				
PR	22-AUG-2000; 2000US-0227182.				

PR	23-AUG-2000	2000US-0227609
PR	30-SEP-2000	2000US-0228982
PR	01-SEP-2000	2000US-0229387
PR	01-SEP-2000	2000US-0229344
PR	01-SEP-2000	2000US-0229345
PR	01-SEP-2000	2000US-0229345
PR	05-SEP-2000	2000US-0229509
PR	05-SEP-2000	2000US-0229637
PR	05-SEP-2000	2000US-0230438
PR	06-SEP-2000	2000US-0230438
PR	08-SEP-2000	2000US-0231442
PR	08-SEP-2000	2000US-0231443
PR	08-SEP-2000	2000US-0231444
PR	08-SEP-2000	2000US-0231445
PR	08-SEP-2000	2000US-0231446
PR	08-SEP-2000	2000US-0231447
PR	08-SEP-2000	2000US-0231448
PR	08-SEP-2000	2000US-0231449
PR	11-SEP-2000	2000US-0231968
PR	14-SEP-2000	2000US-0232397
PR	14-SEP-2000	2000US-0232398
PR	14-SEP-2000	2000US-0232399
PR	14-SEP-2000	2000US-0232400
PR	14-SEP-2000	2000US-0232401
PR	14-SEP-2000	2000US-0233063
PR	14-SEP-2000	2000US-0233065
PR	14-SEP-2000	2000US-0233066
PR	21-SEP-2000	2000US-0233422
PR	21-SEP-2000	2000US-0233423
PR	21-SEP-2000	2000US-0233424
PR	25-SEP-2000	2000US-0234997
PR	25-SEP-2000	2000US-0234998
PR	25-SEP-2000	2000US-0235484
PR	27-SEP-2000	2000US-0235834
PR	27-SEP-2000	2000US-0235836
PR	29-SEP-2000	2000US-0236327
PR	29-SEP-2000	2000US-0236327
PR	29-SEP-2000	2000US-0236356
PR	29-SEP-2000	2000US-0236358
PR	29-SEP-2000	2000US-0236370
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237039
PR	02-OCT-2000	2000US-0237040
PR	13-OCT-2000	2000US-0239393
PR	13-OCT-2000	2000US-0239397
PR	20-OCT-2000	2000US-0241809
PR	20-OCT-2000	2000US-0241816
PR	01-NOV-2000	2000US-0244512
PR	08-NOV-2000	2000US-0246475
PR	08-NOV-2000	2000US-0246476
PR	08-NOV-2000	2000US-0246477
PR	08-NOV-2000	2000US-0246478
PR	08-NOV-2000	2000US-0246479
PR	08-NOV-2000	2000US-0246524
PR	08-NOV-2000	2000US-0246525
PR	08-NOV-2000	2000US-0246526
PR	08-NOV-2000	2000US-0246527
PR	08-NOV-2000	2000US-0246532
PR	08-NOV-2000	2000US-0246532
PR	08-NOV-2000	2000US-0246610
PR	08-NOV-2000	2000US-0246611
PR	08-NOV-2000	2000US-0246613
PR	17-NOV-2000	2000US-0249207
PR	17-NOV-2000	2000US-0249208
PR	17-NOV-2000	2000US-0249209

QY	502	AEEHIIKLCANLAEFTIKTYKEVLDLQKNOIOEPCCKFAIADATREDPFKELIIMRI	561
Db	68	AEEHIIKLCANLAEFTIKTYKEVLDLQKNOIOEPCCKFAIADATREDPFKELIIMRI	127
QY	562	QTYCOMSPMSDFGTOPEQWAIOMEKKAARKGNRKERYCAEHLRYNEALQINLIIMID	621
Db	128	QTYCOMSPMSDFGTOPEQWAIOMEKKAARKGNRKERYCAEHLRYNEALQINLIIMID	187
QY	622	AYTHLEFFTYEEDKRFVAIEDSDSGDDEYCDGDEDEDDLKRLKLEIDREFTMLTF	681
Db	188	AYTHLEFFTYEEDKRFVAIEDSDSGDDEYCDGDEDEDDLKRLKLEIDREFTMLTF	247
QY	682	ENNNMLRLMENPEYENEKLTKLRLNTIMEQYRTBESAGILFTRTOSAVALSQWILEN	741
Db	248	ENNNMLRLMENPEYENEKLTKLRLNTIMEQYRTBESAGILFTRTOSAVALSQWILEN	307
QY	742	EKEFAEYGVKAHHLIGAGHSSEFKPMTQNEQKEYISKFRGKINLLIATTVAHEG:DIKEC	801
Db	308	EKEFAEYGVKAHHLIGAGHSSEFKPMTQNEQKEYISKFRGKINLLIATTVAHEG:DIKEC	365
QY	802	NIVIRGILVNTETIAM-VQANGRAADESTYVLVAHSGVIE	842
Db	366	NVTLKSVWSSPMKYRGPARGRARADESTYVLL-QMSGYID	406
RESULT	6		
ID	AA040129	standard; Protein: 356 AA.	
XX	AA040129;		
XX	22-OCT-2001	(first entry)	
XX			
DE	Human polypeptide SEQ ID NO 3274.		
XX			
KW	Human; neotrophic; immunosuppressant; cytostatic; gene therapy; cancer;		
KW	peripheral nervous system; neuropathy; central nervous system; CNS;		
KW	Alzheimer's; Parkinson's disease; Huntington's disease; hemostatic;		
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;		
KW	chemokine; thrombolytic; drug screening; arthritis; inflammation;		
XX	leukaemia.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200153312-A1.		
XX			
PD	26-JUL-2001.		
XX			
PF	26-DEC-2000; 2000MO-00S34263.		
XX			
PR	21-JAN-2000; 2000US-0488725.		
PR	25-APR-2000; 2000US-0552317.		
PR	09-JUL-2000; 2000US-0598042.		
PR	19-JUL-2000; 2000US-0620312.		
PR	03-AUG-2000; 2000US-0653450.		
PR	14-SEP-2000; 2000US-0662191.		
PR	19-OCT-2000; 2000US-0693036.		
PR	29-NOV-2000; 2000US-0727344.		
XX			
PA	(HYSE-) HYSEQ INC.		
XX			
PI	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D,		
PI	Wang J, Wang Z, Wehman T, Xu C, Xue AJ, Yang Y, Zhan J,		
PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;		
XX			
XX	WPI: 2001-442253/47.		
DR	N-PSDB: AA159285.		
XX			
PT	Novel nucleic acids and polypeptides, useful for treating disorders		
PT	such as central nervous system injuries -		
PS	Example 5; SEQ ID NO 3274; 10078bp; English.		

AAB93708
 ID AAB93708 standard; Protein: 678 AA.
 AC AAB93708;
 DT 26-JUN-2001 (first entry)
 DE Human protein sequence SEQ ID NO:13299.
 KW Human; primer: detection; diagnosis; antisense therapy; gene therapy.
 OS Homo sapiens.
 PN EPI074617-A2.
 PD 07-FEB-2001.
 PE 28-JUL-2000; 2000EP-0116126.
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 PA (HELI-) HELIX RES INST.
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR MPI: 2001-318749/34.
 PT primer sets for synthesizing polynucleotides; particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 PS Claim 8; SEQ ID 13299; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166; AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
 CC AAH93893 represent human amino acid sequences; and AAH15629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SO Sequence 678 AA.
 Query Match 25.38; Score 1344; DR 22; Length 678;
 Best Local Similarity 40.68; Pred. No. 2c-9%;
 Matches 291; Conservative 136; Mismatches 238; Indels 52; Gaps 12;
 QY 306 IOLRPOMEYAPALGKNIICLPGSGKTRAAVYIAKHDKKKAS:PGKVLYLVNK 365
 DB 1 MELRSYOMEYIMPALEGNKNTITMLPGAGKTRAAVAAKHHLE---TVGAVVAVLVNK 56
 QY 366 VLLVEOLFKEQPFLEKKMYRVIGLSGTOLKISPEVVKSCDIIISTA. LENSILNI 425

DB 57 VHLVYQ-AGEEPRMALDGRMTVTTLTSGDMGPRAGCHLARCHDILLCTAELICMTTSP 115
 QY 426 NGEDAGYOLDSFSLIIIDECCHTNKEAYNNINRMTYIMOKLNNKLNKKNKVI 485
 DB 116 --EEHEVELVPSLIVDECHTHRDYNNIMSOYLKLEKLDRAQ-----LPPVL 164
 QY 486 GLTASPGVGATGAKAKAEHILKLCANLDAFTIKVYENLQDLKQIOEPCKK 545
 DB 165 GLTASPGVGATGAKAKAEHILKLCANLDAFTIKVYENLQDLKQIOEPCKK 545
 QY 546 REDPPEKLELIMTRIQTQCSNPS-DFGTQPYQMAIQEKKAKKKNKPKYAEHL 604
 DB 225 SDDPGLDKLKLMDQIHDLKEMPELSRKFGTQYQVVKLSAALAGLQFGKAVYALH 284
 QY 605 RYFNALQINQITRMIDATVHLETYNEEKDKKFAVEDSDSGDDEYCDLE 664
 DB 285 RRYNDALLHDYRAVDLAALADDFYHREHYTKTIL-----C----- 322
 QY 665 KPLKIDETDFLMTLEFFENKKLKLRLAENPEYENKLTIRNTIMEOYTRTESAKGIF 724
 DB 323 -----AERLLALFDKRNKELHLATHGP-ENKLEMLEKTIQROTS-SSNS:FGIIF 373
 QY 725 TKTRQSAVALSQWITENKFAEVGAKHLLGAGHSEFKPTQNEQKQEVISKI 784
 DB 374 TRTRQSAHSLMLMDQOQGLQTVDIRADLLIGAGNSQSTHTHQDQGEVIGK 433
 QY 785 LLIATVAEESGIDKECNVIRGLVTNEMIAMVQARGARADESTYLVANHS:VIEH 844
 DB 434 LLVATVAEESGIDKECNVIRGLVTNEMIAMVQARGARADESTYLVANHS:VIEH 844
 QY 845 TVNDFREKMYKALICVONMKPEYAHKILQOMOSIMEKKKTIRNIAKHKNNSLIT 904
 DB 494 LINEALETMEGAVAAVQMDQAEYQAKRDQALTKRAQAARENRQKFTVEVQ 553
 QY 905 FLCKNSYLACSGEDIHYTEKHHVNMPEFEKELY-IVRENKALOKCADVOIN 963
 DB 554 LKCNQMAVAVGSDLRKVEGTHHVNPNFNSNYNSNDPVYIKVFKDMKPGVISC 613
 QY 964 -CGQAGMTMVIHKGDLPLCLTRNFVVPKNNSTKQYKKVYELDTPPNLDYS 1019
 DB 614 NCGEYVGLQMIYKSVKLPVLKVR--MLLETQGRIOAKKMSRVPFVDFDLQHC 668
 RESULT 9
 ID AAU00296 standard; Protein: 304 AA.
 XX
 AC AAU00296;
 DT 12-SEP-2001 (first entry)
 DE Interferon induced polypeptide, IFN4.
 KW Interferon induced nucleic acid; autoimmune disease; lupus erythematosus;
 KW immunodeficiency; stroke; acquired immunodeficiency syndrome; AIDS;
 KW graft rejection; viral infection; hepatitis; aplastic anaemia; cancer;
 KW human immunodeficiency virus; HIV; immune-mediated glomerulonephritis;
 KW haematologic disease; chronic neutropenia; myocardial infarction;
 KW neurological disease; Alzheimer's disease; Parkinson's disease; tumour;
 KW amyotrophic lateral sclerosis; spinal muscular atrophy; human; IFN4.
 OS Homo sapiens.
 PN WO200118208-A2.
 PD 15-MAR-2001.
 PE 08-SEP-2000; 2000WO-US24704.
 PR 08-SEP-1999; 99US-0152921.
 PR 20-OCT-1999; 99US-0160575.
 PR 20-JAN-2000; 2000US-0177104.

PR 07-SEP-2000; 2000US-0656633.
 XX (CURA-) CURAGEN CORP.
 XX (BIOJ) BIOGEN INC.
 PA
 PI Peyman JA, Da Silva A, Hochman P, Hsu A;
 DR WPI, 2001-235201/24.
 DR N-PSDB; AAS01149.
 PT New interferon induced polypeptides and polynucleotides useful for the
 PT diagnosis, prevention and treatment of immunological, cell
 PT proliferative disorders, such as lupus erythematosus, cancer, stroke
 PT and Alzheimer's disease
 PS
 XX Claim 1; Page 29-32; 134pp; English.
 XX
 CC The sequence represents interferon induced polypeptide, IFN4, IFN
 CC nucleic acids and polypeptides are useful for treating or preventing a
 CC pathology associated with IFN polypeptide in a human. They are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of IFN polypeptide or polynucleotide. IFN nucleic
 CC acids, polypeptides and antibodies are useful for diagnosis, prevention
 CC or treatment of variety of immunological and cell proliferative
 CC disorders, such as autoimmune diseases e.g. lupus erythematosus,
 CC immunodeficiency diseases such as acquired immunodeficiency syndrome
 CC (AIDS), graft rejection, viral infections including hepatitis and human
 CC immunodeficiency virus (HIV), immune-mediated glomerulonephritis,
 CC hematologic diseases such as aplastic anaemia and chronic neutropenia
 CC and cancer. In addition they are also useful for treating or diagnosing
 CC various disorders associated with cell death, including myocardial
 CC infection, stroke, neurological diseases including Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis and spinal muscular
 CC atrophy. IFN nucleic acids and polypeptides are also useful for
 CC identifying interferon-like proteins and interferon agonists, for
 CC screening drugs and compounds which inhibit or enhance IFN activity or
 CC function and as targets for the identification of small molecules that
 CC are immunostimulatory, immunosuppressive, or stimulate or suppress normal
 CC cell or tumour cell growth in mammals, including humans
 CC
 SO Sequence 304 AA;
 XX
 Query Match 20.6%; Score 1093; DB 22; Length 304;
 Best Local Similarity 100.0%; Pred. No. 1,86-76;
 Matches 216; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 41 EVKEIOIQTVAATSGNQMAVELLSTLEKGVNHLGWTREVEALRTGSPPLAARYNNPELT 100
 DB 87 EVKEIOIQTVAATSGNQMAVELLSTLEKGVNHLGWTREVEALRTGSPPLAARYNNPELT 146
 QY 101 DLPSPFENAHDEYIQLNLLOPTLVKLLRVLDKCMEEELLTIEDRNFAAENNNCN 160
 DB 147 DLPSPFENAHDEYIQLNLLOPTLVKLLRVLDKCMEEELLTIEDRNFAAENNNCN 206
 QY 161 EGVRELLAKRIYQKEMFSAFLNVLROTGNNELVOELTSDSESNAELTINLSOVDGQY 220
 DB 207 EGVRELLAKRIYQKEMFSAFLNVLROTGNNELVOELTSDSESNAELTINLSOVDGQY 266
 QY 221 EQLLLSTTVQPNLEKEVMGNNSSSSSPADSSVVS 276
 DB 267 EQLLLSTTVQPNLEKEVMGNNSSSSSPADSSVVS 302
 RESULT 10
 ID AAU00298 standard: Protein; 308 AA.
 XX AAU00298;
 AC AAU00298;
 XX
 DT 12-SEP-2001 (first entry)
 XX
 DE Interferon induced polypeptide, IFN6.
 XX

KW Interferon induced nucleic acid; autoimmune disease; lupus erythematosus;
 KW immunodeficiency; stroke; acquired immunodeficiency syndrome; AIDS;
 KW graft rejection; viral infection; hepatitis; aplastic anaemia; cancer;
 KW human immunodeficiency virus; HIV; immune-mediated glomerulonephritis;
 KW hematologic disease; chronic neutropenia; myocardial infarction;
 KW neurological disease; Alzheimer's disease; Parkinson's disease; tumour;
 KW amyotrophic lateral sclerosis; spinal muscular atrophy; human; IFN6.
 KM
 XX Homo sapiens.
 XX
 PN WO200118208-A2.
 PD 15-MAR-2001.
 XX
 PF 08-SEP-2000; 2000WO-0524704.
 XX
 PR 08-SEP-1999; 99US-0152921.
 PR 20-OCT-1999; 99US-0160575.
 PR 20-JAN-2000; 2000US-0177104.
 PR 07-SEP-2000; 2000US-0656633.
 PA (CURA-) CURAGEN CORP.
 PA (BIOJ) BIOGEN INC.
 PI Peyman JA, Da Silva A, Hochman P, Hsu A;
 DR WPI, 2001-235201/24.
 DR N-PSDB; AAS01151.
 PT New interferon induced polypeptides and polynucleotides useful for the
 PT diagnosis, prevention and treatment of immunological, cell
 PT proliferative disorders, such as lupus erythematosus, cancer, stroke
 PT and Alzheimer's disease
 PS
 XX Claim 1; Page 40-42; 134pp; English.
 XX
 CC The sequence represents interferon induced polypeptide, IFN4, IFN
 CC nucleic acids and polypeptides are useful for treating or preventing a
 CC pathology associated with IFN polypeptide in a human. They are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of IFN polypeptide or polynucleotide. IFN nucleic
 CC acids, polypeptides and antibodies are useful for diagnosis, prevention
 CC or treatment of variety of immunological and cell proliferative
 CC disorders, such as autoimmune diseases e.g. lupus erythematosus,
 CC immunodeficiency diseases such as acquired immunodeficiency syndrome
 CC (AIDS), graft rejection, viral infections including hepatitis and human
 CC immunodeficiency virus (HIV), immune-mediated glomerulonephritis,
 CC hematologic diseases such as aplastic anaemia and chronic neutropenia
 CC and cancer. In addition they are also useful for treating or diagnosing
 CC various disorders associated with cell death, including myocardial
 CC infection, stroke, neurological diseases including Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis and spinal muscular
 CC atrophy. IFN nucleic acids and polypeptides are also useful for
 CC identifying interferon-like proteins and interferon agonists, for
 CC screening drugs and compounds which inhibit or enhance IFN activity or
 CC function and as targets for the identification of small molecules that
 CC are immunostimulatory, immunosuppressive, or stimulate or suppress normal
 CC cell or tumour cell growth in mammals, including humans.
 CC
 SO Sequence 308 AA;
 XX
 Query Match 20.6%; Score 1093; DB 22; Length 308;
 Best Local Similarity 100.0%; Pred. No. 1,96-76;
 Matches 216; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 41 EVKEIOIQTVAATSGNQMAVELLSTLEKGVNHLGWTREVEALRTGSPPLAARYNNPELT 100
 DB 91 EVKEIOIQTVAATSGNQMAVELLSTLEKGVNHLGWTREVEALRTGSPPLAARYNNPELT 150
 QY 101 DLPSPFENAHDEYIQLNLLOPTLVKLLRVLDKCMEEELLTIEDRNFAAENNNCN 160
 DB 151 DLPSPFENAHDEYIQLNLLOPTLVKLLRVLDKCMEEELLTIEDRNFAAENNNCN 210

QY 161 ESGVRELLKRIYQKEMFSAFLNVLROTGNMELVQELTSGDCSESNAEITNLSQVDPGV 220
 |||||||
 DB 211 ESGVRELLKRIYQKEMFSAFLNVLROTGNMELVQELTSGDCSESNAEITNLSQVDPGV 270
 |||||||
 QY 221 EBOULLSTVQPNLEKEVGMENNSSESSPADSSVVS 256
 |||||||
 DB 271 EBOULLSTVQPNLEKEVGMENNSSESSPADSSVVS 306
 |||||||
 RESULT 11
 AAU00297
 ID AAU00297 standard; Protein: 267 AA.
 AC AAU00297;
 XX
 DT 12-SEP-2001 (first entry)
 DE Interferon induced polypeptide, IFN5.
 XX
 KM Interferon induced nucleic acid; autoimmune disease; lupus erythematosus;
 KM immunodeficiency; stroke; acquired immunodeficiency syndrome; AIDS;
 KM graft rejection; viral infection; hepatitis; aplastic anaemia; cancer;
 KM human immunodeficiency virus; HIV; immune-mediated glomerulonephritis;
 KM haematologic disease; chronic neutropenia; myocardial infarction;
 KM neurological disease; Alzheimer's disease; Parkinson's disease; tumour;
 KM amyotrophic lateral sclerosis; spinal muscular atrophy; human; IFN5.
 XX
 OS Homo sapiens.
 XX
 PN WO200118208-A2.
 XX
 PD 15-MAR-2001.
 XX
 PE 08-SEP-2000; 2000MO-US24704.
 XX
 PR 08-SEP-1999; 99US-0152921.
 PR 20-OCT-1999; 99US-0160575.
 PR 20-JAN-2000; 2000US-0177104.
 PR 07-SEP-2000; 2000US-0656633.
 PA (CURA-) CURAGEN CORP.
 PA (BIO) BIOGEN INC.
 PI Peyman JA, Da Silva A, Hochman P, Hsu A;
 DR MPI: 2001-235201/24.
 DR N-PSDB: AAS01150.
 XX
 PT New interferon induced polypeptides and polynucleotides useful for the
 PT diagnosis, prevention and treatment of immunological, cell
 PT proliferative disorders, such as lupus erythematosus, cancer, stroke
 PT and Alzheimer's disease
 XX
 PS Claim 1; Page 33-35; 134pp; English.
 XX
 CC The sequence represents interferon induced polypeptide, IFN5. IFN
 CC nucleic acids and polypeptides are useful for treating or preventing a
 CC pathology associated with IFN polypeptide in a human. They are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of IFN polypeptide or polynucleotide. IFN nucleic
 CC acids, polypeptides and antibodies are useful for diagnosis, prevention
 CC or treatment of variety of immunological and cell proliferative
 CC disorders, such as autoimmune diseases e.g. lupus erythematosus.
 CC immunodeficiency diseases such as acquired immunodeficiency syndrome
 CC (AIDS), graft rejection, viral infections including hepatitis and human
 CC immunodeficiency virus (HIV), immune-mediated glomerulonephritis,
 CC haematologic diseases such as aplastic anaemia and chronic neutropenia
 CC and cancer. In addition they are also useful for treating or diagnosing
 CC various disorders associated with cell death, including myocardial
 CC infarction, stroke, neurological diseases including Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis and spinal muscular
 CC atrophy. IFN nucleic acids and polypeptides are also useful for
 CC identifying interferon-like proteins and interferon agonists, for

CC screening drugs and compounds which inhibit or enhance IFN activity or
 CC function and as targets for the identification of small molecules that
 CC are immunostimulatory, immunosuppressive, or stimulate or suppress normal
 CC cell or tumour cell growth in mammals, including humans.
 XX
 SQ Sequence 267 AA;
 Query Match 16.2%; Score 861; DB 22; Length 267;
 Best Local Similarity 96.6%; Pred. No. 1.6e-58;
 Matches 170; Conservative 2; Mismatches 4; Indels 0 Gaps 0;
 QY 41 EVKEIOIQTAVATSGNMQAVELLSTLEKGVWHLGCTREFVALRTGSPILAARYVNI ELT 100
 |||||||
 DB 87 EVKEIOIQTAVATSGNMQAVELLSTLEKGVWHLGCTREFVALRTGSPILAARYVNI ELT 146
 |||||||
 QY 101 DLPSPFENAHDEYLOLLNLOPTLYDKLLVRDVLDKMBEELLTIEDRRNIAAENNN 160
 |||||||
 DB 147 DLPSPFENAHDEYLOLLNLOPTLYDKLLVRDVLDKMBEELLTIEDRRNIAAENNN 206
 |||||||
 QY 161 ESGVRELLKRIYQKEMFSAFLNVLROTGNMELVQELTSGDCSESNAEITNLSQV 216
 |||||||
 DB 207 ESGVRELLKRIYQKEMFSAFLNVLROTGNMELVQELTSGDCSESNAEITNLSQV 262
 |||||||
 RESULT 12
 AAU00299
 ID AAU00299 standard; Protein: 271 AA.
 AC AAU00299;
 XX
 DT 12-SEP-2001 (first entry)
 DE Interferon induced polypeptide, IFN7.
 XX
 KM Interferon induced nucleic acid; autoimmune disease; lupus erythematosus;
 KM immunodeficiency; stroke; acquired immunodeficiency syndrome; AIDS;
 KM graft rejection; viral infection; hepatitis; aplastic anaemia; cancer;
 KM human immunodeficiency virus; HIV; immune-mediated glomerulonephritis;
 KM haematologic disease; chronic neutropenia; myocardial infarction;
 KM neurological disease; Alzheimer's disease; Parkinson's disease; tumour;
 KM amyotrophic lateral sclerosis; spinal muscular atrophy; human; IFN7.
 XX
 OS Homo sapiens.
 XX
 PN WO200118208-A2.
 XX
 PD 15-MAR-2001.
 XX
 PE 08-SEP-2000; 2000MO-US24704.
 XX
 PR 08-SEP-1999; 99US-0152921.
 PR 20-OCT-1999; 99US-0160575.
 PR 20-JAN-2000; 2000US-0177104.
 PR 07-SEP-2000; 2000US-0656633.
 PA (CURA-) CURAGEN CORP.
 PA (BIO) BIOGEN INC.
 PI Peyman JA, Da Silva A, Hochman P, Hsu A;
 DR MPI: 2001-235201/24.
 DR N-PSDB: AAS01152.
 XX
 PT New interferon induced polypeptides and polynucleotides useful for the
 PT diagnosis, prevention and treatment of immunological, cell
 PT proliferative disorders, such as lupus erythematosus, cancer, stroke
 PT and Alzheimer's disease
 XX
 PS Claim 1; Page 44-46; 134pp; English.
 XX
 CC The sequence represents interferon induced polypeptide, IFN7. IFN
 CC nucleic acids and polypeptides are useful for treating or preventing a
 CC pathology associated with IFN polypeptide in a human. They are useful for

CC determining the presence of or predisposition to a disease associated
CC with altered levels of IFN polypeptide or polynucleotide. IFN nucleic
CC acids, polypeptides and antibodies are useful for diagnosis, prevention
CC or treatment of variety of immunological and cell proliferative
CC disorders, such as autoimmune diseases e.g. lupus erythematosus,
CC immunodeficiency diseases such as acquired immunodeficiency syndrome
CC (AIDS), graft rejection, viral infections including hepatitis and human
CC immunodeficiency virus (HIV), immune-mediated glomerulonephritis,
CC hematologic diseases such as aplastic anemia and chronic neutropenia
CC and cancer. In addition they are also useful for treating or diagnosing
CC various disorders associated with cell death, including myocardial
CC infarction, stroke, neurological diseases including Alzheimer's and
CC Parkinson's diseases, amyotrophic lateral sclerosis and spinal muscular
CC atrophy. IFN nucleic acids and polypeptides are also useful for
CC identifying interferon-like proteins and interferon antagonists, for
CC screening drugs and compounds which inhibit or enhance IFN activity or
CC function and as targets for the identification of small molecules that
CC are immunostimulatory, immunosuppressive, or stimulate or suppress normal
CC cell or tumour cell growth in mammals, including humans.

XX Sequence 271 AA:

Query Match 16.2% Score 861; DB 22; Length 271;
Best Local Similarity 96.6%; Pred. No. 1,76,58;
Matches 170; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 41 EYKEQIORTVATSGNMQAVELLSTLEKGVMLGWTREVEANIRGSGPIAARYNPBELT 100
DB 91 EYKEQIORTVATSGNMQAVELLSTLEKGVMLGWTREVEANIRGSGPIAARYNPBELT 150
QY 101 DLPSPFENAHDEYLOLLNLQPTLVKLLVQVLDKCHEEELTTEENKPAANNNN 160
DB 151 DLPSPFENAHDEYLOLLNLQPTLVKLLVQVLDKCHEEELTTEENKPAANNNN 210
QY 161 ESGVRELLKRIYCKENMFSAFLNVLKQTNNELVQELTSCDSSESNAETINISQVD 216
DB 211 ESGVRELLKRIYCKENMFSAFLNVLKQTNNELVQELTSCDSSESNAETINISQVD 266

RESULT 13
AAU23654
ID AAU23654 standard; Protein; 448 AA.
XX
AC AAU23654;
DT 18-DEC-2001 (first entry)
XX
DE Novel human enzyme polypeptide #740.
XX
KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW nephrotropic; anticoagulant.
XX
OS Homo sapiens.
XX
PN WO200155301-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01239.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.

PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0244674.
 PR 08-NOV-2000; 2000US-0244675.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI: 2001-465566/50.
 DR N-PSDB: AAS41524.
 XX
 PT Novel polypeptides and polynucleotides useful for diagnosing,
 PT preventing, treating neural, immune system, muscular, reproductive,
 PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
 PT diseases -
 XX
 PS Claim 11: SEQ ID No 1650; 1180pp; English.
 XX
 CC The present invention relates to the isolation of novel human enzyme
 CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences
 CC encoding them. The enzyme polypeptides of the invention may comprise the
 CC functional classes of oxidoreductases, transferases, hydrolases, lyases,

CC isomerases or ligases. The sequences of the invention are useful in the
 CC diagnosis, treatment, prevention and/or prognosis of a wide range of
 CC disorders including hyperproliferative disorders (e.g. cancer),
 CC Immunodeficiency disorders (e.g. AIDS) autoimmune disorders
 CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
 CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
 CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
 CC blood-related disorders (e.g. haemophilia), reproductive disorders
 CC (e.g. infertility) and infectious disorders (e.g. influenza). The
 CC polynucleotides of the invention can also be used in gene therapy.
 CC AAO22915-AA023814 represent the novel human enzyme po-ypeptides of the
 CC invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 448 AA:
 Query Match 15.2%; Score 808; DB 22; Length 418;
 Best Local Similarity 42.4%; Pred. No. 4,9e-54;
 Matches 172; Conservative 76; Mismatches 126; Indels 7; Gaps 5;
 QY 480 PLPOLIGLTPSPVGGATKQAKAEERHLKCANLDATIKTVKENTDQKNKQICKKRF 539
 DB 11 PLPOVIGLTPSPVGGATKQAKAEERHLKCANLDATIKTVKENTDQKNKQICKKRF 539
 QY 540 AIAADATREDPREKLEIMTRIQTCQSPMS-DGPTQPYQMAQMKKAKKQKKER 598
 DB 71 NLCHRRSDPEGDLKLMDOHDHLEMPBLSRKGTQMYEDQYVXKLEAALADIDQR 130
 QY 599 VCAEHARKYNEALQINDIRIMIDATYHLETFYNEEKDKKFAVEDDSDGDDDEVYQDE 658
 DB 131 VYALMLRRYNOLLHDIVRAVDALALADDFYHRRHYKTKQL-----C----- 174
 QY 659 DEDDLKPKLDEIDRFLMTLFFENNNKMKRLAENPEYENKTKLNTIMEQYTKIES 718
 DB 175 -----AERRILALFPDDRKNELALATGPR-ENKRLLEKILQROFSSNS 219
 QY 719 ARGILFTKROSAYVALSOMITENEFKFAEYGVAAHLLTAGSSSEKPK-CNCHQEVISKF 778
 DB 220 PGGILFTKROSASHLLMLDQOQGLQTVDIRAQLLISAGNSOSTHW-TORUCQVIOKRF 279
 QY 779 RTGKINLIATVAEGDIDKECNIVIRGLVTNEIAVQAGRADESTYVIVANSGS 838
 DB 280 QGCTLNLIATVAEEGLDIPHCNVVYRGTLTINISVQAGRAKXAOQSVYAVATGGS 339
 QY 839 GVIEHETVNDERKMMYKRAHCQVNMKPEYAKHKLLEQOMOSIMEK 884
 DB 340 RELKRELINLELETLMEOVAVAVKMDQAEYQAKIRDLQAAALTRR 385
 RESULT 14
 ABB42219
 ID ABB42219 standard; Peptide; 166 AA.
 XX
 AC ABB42219;
 XX
 DT 04-FEB-2002 (first entry)
 XX
 DE Peptide #9725 encoded by human foetal liver single exon probe.
 XX
 KW Human; foetal liver; gene expression; single exon nucleic acid probe.
 XX
 OS Homo sapiens.
 XX
 PN WO200157277-A2.
 XX
 PD 09-AUG-2001.
 XX
 XX 30-JAN-2001; 2001WO-US00669.
 PF 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.

```

PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human fetal liver.
XX
XX Claim 27; SEQ ID NO 34854; 639pp + sequence listing; English.
XX
XX The invention relates to a single exon nucleic acid probe for
XX measuring human gene expression in a sample derived from human foetal
XX liver. The single exon nucleic acid probes may be used for predicting,
XX measuring and displaying gene expression in samples derived from human
XX foetal liver. The present sequence is a peptide encoded by a single exon
XX nucleic acid probe of the invention.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pat_sequences.
XX
XX Sequence 166 AA:
SQ
Query Match 14.7%: Score 783; DB 22; Length 166;
Best Local Similarity 93.3%: Pred. No. 9.1e-53;
Matches 152; Conservative 3; Mismatches 8; Indels 0; Gaps 0;
OY 1 MSGNSTDENFRYLISCFRARKMYIOVEPVLDYTLFLPAEYKEQIORTVAISGNQAVE 60
DB 1 MSGYSTDENFRYLISCFRARKMYIOVEPVLDYTLFLPAEYKEQIORTVAISGNQAVE 60
OY 61 LLSTLEKGVNHLGWTREVEALRRTGSPPLAARYNMPETDLPSPSFENAHDEYJLLNL 120
DB 61 LLSTLEKGVNHLGWTREVEALRRTGSPPLAARYNMPETDLPSPSFENAHDEYJLLNL 120
OY 121 LQPTLVDKLLVRDVLDKCMEELLTIEDNRNIAAENNGNSG 163
DB 121 LQPTLVDKLLVRDVLDKCMEELLTIEDNRNRYGVCSGDASLLG 163
RESULT 15
AAM63104
ID AAM63104 strand: Protein: 166 AA.
XX
XX AAM63104;
XX
XX 05-NOV-2001 (first entry)
XX
XX Human brain expressed single exon probe encoded protein SEQ ID NO: 35209.
XX
XX Human; brain expressed exon: gene expression analysis: probe;
XX microarray; Alzheimer's disease; multiple sclerosis; s. Alzheimer's;
XX epilepsy; cancer.
XX
XX Homo sapiens.
XX
XX WO200157275-A2.
XX
XX 09-AUG-2001.
XX
XX 30-JAN-2001; 2001WO-US00667.
XX
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.

```

```

PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483446/52.
XX
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains.
XX
XX Example 4; SEQ ID NO: 35209; 650pp + Sequence listing; English.
XX
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, s. Alzheimer's;
XX epilepsy and cancers. The present sequence is a protein encoded by one of
XX the probes of the invention.
XX
XX Sequence 166 AA:
SQ
Query Match 14.7%: Score 783; DB 22; Length 166;
Best Local Similarity 93.3%: Pred. No. 9.1e-53;
Matches 152; Conservative 3; Mismatches 8; Indels 0; Gaps 0;
OY 1 MSGNSTDENFRYLISCFRARKMYIOVEPVLDYTLFLPAEYKEQIORTVAISGNQAVE 60
DB 1 MSGYSTDENFRYLISCFRARKMYIOVEPVLDYTLFLPAEYKEQIORTVAISGNQAVE 60
OY 61 LLSTLEKGVNHLGWTREVEALRRTGSPPLAARYNMPETDLPSPSFENAHDEYJLLNL 120
DB 61 LLSTLEKGVNHLGWTREVEALRRTGSPPLAARYNMPETDLPSPSFENAHDEYJLLNL 120
OY 121 LQPTLVDKLLVRDVLDKCMEELLTIEDNRNIAAENNGNSG 163
DB 121 LQPTLVDKLLVRDVLDKCMEELLTIEDNRNRYGVCSGDASLLG 163

```

Search completed: May 8, 2003, 15:52:39
Job time : 86 secs

QY 2795 TACAGATGCAAGATTAATGAGAAAAAGAAAAAGCAAGAGAAATATTCACAGACTT 2854
 |||||||
 DB 242 TACAGATGCAAGATTAATGAGAAAAAGAAAAAGCAAGAGAAATATTCACAGACTT 301
 QY 2855 ACAAGAAATACCCATCCTAATAACTTCTTGGAAAAAGAGAGTGTATAGCTGT 2914
 |||||||
 DB 302 ACAAGAAATACCCATCCTAATAACTTCTTGGAAAAAGAGAGTGTATAGCTGT 361
 QY 2915 CTGGGAGAGATATTCATTAATGAGAAAAATGCATCAGCTCATATGAGTACAGATTC 2974
 |||||||
 DB 362 CTGGGAGAGATATTCATTAATGAGAAAAATGCATCAGCTCATATGAGTACAGATTC 421
 QY 2975 AGGAACCTTACATTTGTAAGAGAAAAACAAGCAGTCAAAAGAGTGTCTACTATCAA 3034
 |||||||
 DB 422 AGGAACCTTACATTTGTAAGAGAAAAACAAGCAGTCAAAAGAGTGTCTACTATCAA 481
 QY 3035 TAAATGAGAAATCATCTGCAAAATGTGGCAGGCTTGGAGCAATGATCTGCAAG 3094
 |||||||
 DB 482 TAAATGAGAAATCATCTGCAAAATGTGGCAGGCTTGGAGCAATGATCTGCAAG 541
 QY 3095 GCTTAGATTGCTGCTGCTGCAAAATAGCAATTTGAGTGTTCATTAATATTC 3154
 |||||||
 DB 542 GCTTAGATTGCTGCTGCTGCTGCAAAATAGCAATTTGAGTGTTCATTAATATTC 601
 QY 3155 CAAGAAACATACAAAAAGTGGTAGAATTTACTATCAGATTCCTCAATCTTGCATAT 3214
 |||||||
 DB 602 CAAGAAACATACAAAAAGTGGTAGAATTTACTATCAGATTCCTCAATCTTGCATAT 661
 QY 3215 CAGAAATGCTTTTATTTAGTATGAGAGATTTAGCTTATTTAGATCTT TTAATAATG 3274
 |||||||
 DB 662 CAGAAATGCTTTTATTTAGTATGAGAGATTTAGCTTATTTAGATCTT TTAATAATG 721
 QY 3275 TATCAGTTAAACATTAATATGATATGATATGATATGATATGATATGATATG 3334
 |||||||
 DB 722 TATCAGTTAAACATTAATATGATATGATATGATATGATATGATATGATATG 781
 QY 3335 TAAGAAATCAATAAATGATTTGTTTACTCTG 3365
 |||||||
 DB 782 TAAGAAATCAATAAATGATTTGTTTACTCTG 812

RESULT 3
 LOCUS B0960157 870 bp mRNA 1 line EST 21-AUG-2002
 DEFINITION AGEMCOURT_8949486 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6472264
 5', mRNA sequence.

ACCESSION B0960157
 VERSION B0960157.1 GI:22375635
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 870)
 AUTHORS NIH-MGC http://mgi.nhl.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@dbp-remail.nih.gov
 Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L14M14005 row: 0 column: 17
 High quality sequence stop: 738.

FEATURES
 source 1..870
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:6472264"

BASE COUNT 319 a 131 c 204 g 214 t 2 others
 ORIGIN
 Query Match 23.8% Score 801.8; DB 14; Length 870;
 Best Local Similarity 98.4%; Pred. No. 2.3e-161;
 Matches 851; Conservative 0; Mismatches 9; Indels 5; Gaps 4;
 QY 1940 AAAAAAGAAATGCAAAAGAGCTGTTTGGCAAGCAATTTGAGAGATACAGAGCC 1999
 |||||||
 DB 1 AAGAAAGAAATGCAAAAGAGCTGTTTGGCAAGCAATTTGAGAGATACAGAGCC 59
 QY 2000 TACAATTAATGACCAATTTGAAATGATAGATGGATATCTGATCTTGAAT 2059
 |||||||
 DB 60 TACAATTAATGACCAATTTGAAATGATAGATGGATATCTGATCTTGAAT 119
 QY 2060 ATGAAGAGAAATGATAGAGTTTGGATGATAGAGATGATAGATGATAGAGGCTG 2119
 |||||||
 DB 120 ATGAAGAGAAATGATAGAGTTTGGATGATAGAGATGATAGATGATAGAGGCTG 179
 QY 2120 ATGAGTATTTGATGATGATAGATGATAGATGATGATGATGATGATGATG 2179
 |||||||
 DB 180 ATGAGTATTTGATGATGATAGATGATGATGATGATGATGATGATGATG 239
 QY 2180 AAGCAGATAGATTTGATGATGATGATGATGATGATGATGATGATGATGATG 2239
 |||||||
 DB 240 AAGCAGATAGATTTGATGATGATGATGATGATGATGATGATGATGATGATG 229
 QY 2240 CTGAAGACCCGGAATGATGATGATGATGATGATGATGATGATGATGATGATG 2299
 |||||||
 DB 300 CTGAAGACCCGGAATGATGATGATGATGATGATGATGATGATGATGATGATG 359
 QY 2300 AATATCTAGAGCTGAGAGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2359
 |||||||
 DB 360 AATATCTAGAGCTGAGAGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 419
 QY 2360 CATATGCGCTTCCAGAGAGATTTACTGAAATGAAAAATTTGCTGATAGAGAG 2419
 |||||||
 DB 420 CATATGCGCTTCCAGAGAGATTTACTGAAATGAAAAATTTGCTGATAGAGAG 479
 QY 2420 CCCACCATCTGATTTGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2479
 |||||||
 DB 480 CCCACCATCTGATTTGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 539
 QY 2480 AAAAAAGCTATTAATTAATTTGCACTGCAAAATCAATCTGCTATACAGAG 2539
 |||||||
 DB 540 AAAAAAGCTATTAATTAATTTGCACTGCAAAATCAATCTGCTATACAGAG 599
 QY 2540 TGGCAGAGAAAGCTGATGATTAATAAGATGATACATTTGATGCTTATGCTCA 2599
 |||||||
 DB 600 TGGCAGAGAAAGCTGATGATTAATAAGATGATACATTTGATGCTTATGCTCA 659
 QY 2600 CCAATGAATATGACCTGATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2659
 |||||||
 DB 660 CCAATGAATATGACCTGATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 719
 QY 2660 TCGTGTGCTGACAGT-GGTGAGAGATTAAGCAATGAGAGAGAGAGAGAGAG 2718
 |||||||
 DB 720 TCGTGTGCTGACAGTGTGAGAGATTAAGCAATGAGAGAGAGAGAGAGAG 779
 QY 2719 GAGAGATGATGATTAAGCAATGATGATGATGATGATGATGATGATGATGAT 2775
 |||||||
 DB 780 GAGAGATGATGATTAAGCAATGATGATGATGATGATGATGATGATGATGAT 839
 QY 2776 GCTCATAGATTTTGAATTAAGCA 2800
 |||||||
 DB 840 GCTCATAGATTTTGAATTAAGCA 864

/db_xref="taxon:9606"
 /clone="IMAGE:4183126"
 /clone_lib="NCI CGAP Brn64"
 /tissue_type="gliblastoma with EGFR amplification"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: brain; Vector: pCMV-Sport6; Site: 1; NotI;
 Site: 2; SalI; Cloned unidirectionally; Primer: Cligo dt.
 Average insert size 1.57 kb. Constructed by Life
 Technologies. Note: this is a NCI CGAP library."
 BASE COUNT 353 a 221 c 243 g 218 t
 ORIGIN

Query Match 21.1%; Score 711.6; DB 12; Length 1035;
 Best Local Similarity 97.2%; Pred. No. 5,3e-142;
 Matches 756; Conservative 0; Mismatches 19; Indels 3; Gaps 3;

QY 1049 AGAATGTGCGAGAGAGAGATCCCGGAGCCAGCACTCAGTCCGCT ACCAAATGG 1108
 DB 4 AGAATGTGCGAGAGAGATCCCGGAGCCAGCACTCAGTCCGCT ACCAAATGG 63
 QY 1109 AAGTCCCGAGCGAGCTTGGAGGAGATATATATATCTGCGCTCCCTAGGAGATG 1168
 DB 64 AAGTCCCGAGCGAGCTTGGAGGAGATATATATATCTGCGCTCCCTAGGAGATG 123
 QY 1169 GAAAAACGAGTGCCTGTTTACATTCGCAAGATCCTAGACAGAA AAAAAAGCAT 1228
 DB 124 GAAAAACGAGTGCCTGTTTACATTCGCAAGATCCTAGACAGAA AAAAAAGCAT 183
 QY 1229 CTCAGCTGGAAAGTTATAGTCTTGTCAATAGTACTGCTAGTTGAATAGCTTC 1288
 DB 184 CTCAGCTGGAAAGTTATAGTCTTGTCAATAGTACTGCTAGTTGAATAGCTTC 243
 QY 1289 GCAAGAGTTCACACATTTTGAAGAAATGTGATGCTTATGATTAATGCTGATGA 1348
 DB 244 GCAAGAGTTCACACATTTTGAAGAAATGTGATGCTTATGATTAATGCTGATGA 303
 QY 1349 CCCAATCGAAATATATATATTCGAGATGTCAAGTCTGCTGATATATATATACAGTACG 1408
 DB 304 CCCAATCGAAATATATATATTCGAGATGTCAAGTCTGCTGATATATATATACAGTACG 363
 QY 1409 CTCGAATCTTGGAAACTCCCTCTTAACCTTGAAGAAATGAGAGATGCTGCTGAT 1468
 DB 364 CTCGAATCTTGGAAACTCCCTCTTAACCTTGAAGAAATGAGAGATGCTGCTGAT 423
 QY 1469 TGTCAACTTTTCCCTCATATATGATGATGATGATGATGATGATGATGATGATGAT 1528
 DB 424 TGTCAACTTTTCCCTCATATATGATGATGATGATGATGATGATGATGATGATGAT 483
 QY 1529 ATATATATCATGATGAGCATATATATGATGATGATGATGATGATGATGATGATGAT 1588
 DB 484 ATATATATCATGATGAGCATATATATGATGATGATGATGATGATGATGATGATGAT 543
 QY 1589 AAAACAACGATGATGCTCCCTCTCAAGATGATGATGATGATGATGATGATGATGAT 1648
 DB 544 AAAACAACGATGATGCTCCCTCTCAAGATGATGATGATGATGATGATGATGATGAT 603
 QY 1649 GAGGGCGAG 1708
 DB 604 GAGGGCGAG 662
 QY 1709 ATGCAATTTCTATTAAGCTGTTAAGAAAGCTTGTATCACTGAAAGAAATCAATCAAG 1768
 DB 663 ATGCAATTTCTATTAAGCTGTTAAGAAAGCTTGTATCACTGAAAGAAATCAATCAAG 722
 QY 1769 AGCAGATGAG 1826
 DB 723 GGCATGCGAG 778

RESULT 6 BF983236 729 bp mRNA line:1 EST 23-JAN-2001
 LOCUS BF983236
 DEFINITION 602305873F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4397083 5',

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 NIH-MGC <http://mgs.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT Robert Strausberg, Ph.D.
 Email: rsraus@mail.nih.gov
 Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LIML at:
<http://image.liml.gov>
 Plate: LIML0097 row: a column: 20
 High quality sequence stop: 665.

FEATURES
 source
 Location/Qualifiers

1..729
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4397083"
 /clone_lib="NIH_MGC_88"
 /tissue_type="duodenal adenocarcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: small intestine; Vector: pCMV-Sport6;
 Site: 1; NotI; Site: 2; SalI; Cloned unidirectionally;
 oligo-dt primed. Average insert size 1.767 kb. Library
 enriched for full-length clones and constructed by Life
 Technologies. Note: this is a NIH_MGC Library."
 BASE COUNT 240 a 137 c 184 g 168 t
 ORIGIN

Query Match 20.4%; Score 687; DB 12; Length 729;
 Best Local Similarity 98.9%; Pred. No. 9.2e-137;
 Matches 723; Conservative 0; Mismatches 5; Indels 3; Gaps 3;

QY 586 GAGGAGAACTGTGCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 645
 DB 1 GAGGAGAACTGTGCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
 QY 646 AATGATCAGTGTGAG 705
 DB 61 AATGATCAGTGTGAG 120
 QY 706 GCATTTCTCAATGTTCTTCTGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 765
 DB 121 GCATTTCTCAATGTTCTTCTGCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
 QY 766 TGTGATGCTGAG 825
 DB 181 TGTGATGCTGAG 240
 QY 826 GTGGAAGAGCACTCTTTCACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 885
 DB 241 GTGGAAGAGCACTCTTTCACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
 QY 886 ATGAGAGATTAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 945
 DB 301 ATGAGAGATTAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 360
 QY 946 ACAGAGTTGGAG 1005
 DB 361 ACAGAGTTGGAG 420
 QY 1006 ATGGGAGATGATTCAGGAG 1065

Db 421 ATGGGAGATGATTCAGGACGACATGGGAAGTATTGATGAGAGA:AAATCTGGCAGACAGA 480
 Oy 1066 GCATCCCGGAGCCAGAACTCCAGCTCAGGCGCTTACCAATGAGAGTTG:CGAGCCAGCC 1125
 Db 481 GCATCCCGGAGCCAGAACTCCAGCTCAGGCGCTTACCAATGAGAGTTG:CGAGCCAG 539
 Oy 1126 TTGGAGGAGAAATATCATCATCTGCTCCCTCAGGAGGAGTGGAAAAA:AGAGTGGCT 1185
 Db 540 TTGGAGGAGAAATATCATCATCTGCTCCCTCAGGAGGAGTGGAAAAA:AGAGTGGCT 598
 Oy 1186 GTTACATTTGCCAGGATCAGTACAGCAAGAAAAAGCATCTGAGG:GGAAAGCTT 1245
 Db 599 GTTACATTTGCCAGGATCAGTACAGCAAGAAAAAGCATCTGAGG:GGAAAGCTT 657
 Oy 1246 ATAGTCTTGTCAATAGGACTGCTAGTGAACAGCTCTCCGCAAG:GTTTCCAGCA 1305
 Db 658 ATAGTCTTGTCAATAGGACTGCTAGTGAACAGCTCTCCGCAAG:GTTTCCAGCA 717
 Oy 1306 TTTTTCAGAGA 1316
 Db 718 TTTTTCAGAGA 728
 RESULT 7
 LOCUS BG741146 827 bp mRNA linear EST 15-MAY-2001
 DEFINITION 602631817F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4776869 5',
 mRNA sequence.
 ACCESSION BG741146
 VERSION BG741146.1 GI:14051799
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 827)
 NIH-MGC http://mgs.nci.nih.gov/.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bcrf-remail.nih.gov
 Tissue Procurement: James Cleaver, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA
 Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLM10630 row: b column: 06
 High quality sequence stop: 763.
 FEATURES
 source location/Qualifiers
 1..827
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4776869"
 /clone_id="NCI_CGAP_Skn3"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: skin; Vector: pCMV-SPORT; Site: 1; Nccl;
 Site-2: SalI; Cloned unidirectionally; primer: oligo dT.
 Average insert size 1.5kb. Library constructed by Life
 Technologies. Note: this is a NCI_CGAP library."
 BASE COUNT 254 a 186 c 204 g 183 t
 ORIGIN
 Query Match 19.4%; Score 652.8; DB 12; Length 827;
 Best local similarity 96.1%; Pred. No. 2.1e-129;
 Matches 798; Conservative 0; Mismatches 17; Indels 15; Gaps 12:
 Oy 426 CGGACGCTCTGGCGCGCCGCTACATGAGACCTGAGCTACGCACTTGTCTTCATC 485
 Db 1 CGGACGCTCTGGCGCGCGCCGCTACATGAGACCTGAGCTACGCACTTGTCTTCATC 59
 Oy 486 GTTTCAGAGCGTATGATGATATCTCCAACTGCTGAAGCTCTTCAGTCTCACTGG 544

Db 60 GTTTCAGAGCGTATGATGATATCTCCAACTGCTGAAGCTCTTCAGTCTCACTGG 119
 Oy 545 TGGACAACTTCTAGTTAGAGAGCTGCTGATTAATGATGAGAGAACTGTTGAGACA 604
 Db 120 TGGACAACTTCTAGTTAGAGAGCTGCTGATTAATGATGAGAGAACTGTTGAGACA 179
 Oy 605 TTGAAGACAGAAACCGGATGCTGCTGAGAAAACAAATGAAATGAAATCACTGTAAGAG 664
 Db 180 TTGAAGACAGAAACCGGATGCTGCTGAGAAAACAAATGAAATGAAATCACTGTAAGAG 239
 Oy 665 AGCTACTAAAAAGATTTGTGCAAGAAAAGTGGTTCTCTGATTTTGTGATTTCTTC 724
 Db 240 AGCTACTAAAAAGATTTGTGCAAGAAAAGTGGTTCTCTGATTTTGTGATTTCTTC 299
 Oy 725 GTCAACAGGAAACAAATGAACTTGTCAAGAGTTAAAGAGCTGATGTTG--CTTGAAG 782
 Db 300 GTCAACAGGAAACAAATGAACTTGTCAAGAGTTAAAGAGCTGATGTTGTTGCTCAAAAAG 359
 Oy 783 CAATGAGAGATTTAGAAATTTATCAAGATTGATGCTCTCAAGTGAAGACAACTTCT 842
 Db 360 CAATGAGAGATTTAGAAATTTATCAAGATTGATGCTCTCAAGTGAAGACAACTTCT 419
 Oy 843 TTCAACAGAGTTGAGCAAAATCTGGA--GAAGAGAGTGTGGGGATGGAGAAATCACTGAT 901
 Db 420 TTCAACAGAGTTGAGCAAAATCTGAGAAAGAGTGTGGGGATGGAGAAATCACTGAT 479
 Oy 902 CAGAAATCATCTTTTGA--GATTTCTCTAGTTTTCAGATGAGACAAAGTTTGA--G 958
 Db 480 CAGAAATCATCTTTTGA--GATTTCTCTAGTTTTCAGATGAGACAAAGTTTGA--G 539
 Oy 959 AAGGAAGTGTCAAGCTGCTTACATGAAAGTCTTGA--CATTAACAGACATGGCAGTGTAT 1017
 Db 540 AAGGAAGTGTCAAGCTGCTTACATGAAAGTCTTGA--CATTAACAGACATGGCAGTGTAT 599
 Oy 1018 TCAGGACACATGGGAAGTATGATTCAGATGAA--GAGAAATGGCAGCAAGATCTCCGGA 1076
 Db 600 TCAGGACACATGGGAAGTATGATTCAGATGAA--GAGAAATGGCAGCAAGATCTCCGGA 659
 Oy 1077 GCCAAGACTCCAGCTCAGGCGCTTACCAAAATG--AAGTTGCCAGCCAGCCCTGGAAGG 1134
 Db 660 GCCAAGACTCCAGCTCAGGCGCTTACCAAAATGGAAGTGTGCGCAGCCAG--CTTGACAGG 718
 Oy 1135 AAGAAATATCATCTGCTGCTGCTTACAGAGAGTGGAAAAACCAAGTGGCTTTACATT 1194
 Db 719 AAGAAATATCATCTGCTGCTGCTTACAGAGAGTGGAAAAACCAAGTGGCTTTACATT 777
 Oy 1195 -GCCAAGATCAGTACAGCAAGAAAAAAGCATGAGCTGAGAAAG 124
 Db 778 AGCCAGGATCAGTACAGCAAGAAAAAAGCATGAGCTGAGAAAG 827
 RESULT 8
 LOCUS BF686405 781 bp mRNA linear EST 22-DEC-2000
 DEFINITION 602144786F1 NIH-MGC_46 Homo sapiens cDNA clone IMAGE:4304805 5',
 mRNA sequence.
 ACCESSION BF686405
 VERSION BF686405.1 GI:11971813
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 781)
 NIH-MGC http://mgs.nci.nih.gov/.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cga@bcrf-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory;
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1CML170 row: 1 column: 22
High quality sequence stop: 731.

FEATURES
source
location/Qualifiers

1..781
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4304805"
/clone_lib="NIH_MGC_46"
/tissue_type="Telomysarcoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCR8.1; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dt priming. Additionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

BASE COUNT 235 a 168 c 204 g 174 t
ORIGIN

Query Match 19.4%; Score 651.8; DB 12; Length 781;
Best Local Similarity 99.4%; Pred. No. 3.3e-129;

Matches 675; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 385 CTGTGGTGGACTGGGAATTCGTGGAGGCCCTCCGAGAACCGGAGCCTTTGGCGGCC 444

Db 5 CCTGTGGGACCTGGGAATTCGTGGAGGCCCTCCGAGAACCGGAGCCTTTGGCGGCC 64

QY 445 CGGTATATGAACCTTACGCTGAGGACTGGCTCTCCATCGTTTGAAGATGCTCATGAT 504

Db 65 CGGTATATGAACCTTACGCTGAGGACTGGCTCTCCATCGTTTGAAGATGCTCATGAT 124

QY 505 GAATATCTCCAACTGCTGAACCTCTTCAAGCCCACTCTGTGGAGCAAGCTTTAGTTAGA 564

Db 125 GAATATCTCCAACTGCTGAACCTCTTCAAGCCCACTCTGTGGAGCAAGCTTTAGTTAGA 184

QY 565 GAGCTTGTGATTAAGTATGATGAGAGGAACTGTGACAAATGAGACAAACCGGAT 624

Db 185 GAGCTTGTGATTAAGTATGATGAGAGGAACTGTGACAAATGAGACAAACCGGAT 243

QY 625 GCTGTGTCAGAAAACATGAATGAATCAAGTGTAAAGAGCTTAAATAGAGAT 684

Db 244 GCTGTGTCAGAAAACATGAATGAATCAAGTGTAAAGAGCTTAAATAGAGAT 303

QY 685 CAGAAAGAAACCTGTTCTGATTTGTAATGTTCTTGTCAACAGTAACATGAA 744

Db 304 CAGAAAGAAACCTGTTCTGATTTGTAATGTTCTTGTCAACAGTAACATGAA 363

QY 745 CTGTGCAAGAGTTAAGAGCTCTGATTTGTCAGAAAGCAATGCAGAGAT 804

Db 364 CTGTGCAAGAGTTAAGAGCTCTGATTTGTCAGAAAGCAATGCAGAGAT 423

QY 805 TCACAAAGTTAGTGTCTCAAGTGAAGAGCAACTCTTCAACACAGAT 864

Db 424 TCACAAAGTTAGTGTCTCAAGTGAAGAGCAACTCTTCAACACAGAT 483

QY 865 CTGGAGAGAGGCTGGGGCAGTGAATACTCATCGAATCATCTT 924

Db 484 CTGGAGAGAGGCTGGGGCAGTGAATACTCATCGAATCATCTT 543

QY 925 TCTGTAGTTTCAAGATCAGACAGATTTGGCAGAGAGAGTGTACGCTTTAGATGAA 984

Db 544 TCTGTAGTTTCAAGATCAGACAGATTTGGCAGAGAGAGTGTACGCTTTAGATGAA 603

QY 985 AGCTCTTGACATTAACAG-CACATGGCGACGATTCAGGACCATGGAGATGATTCA 1043

Db 604 AGCTCTTGACATTAACAG-CACATGGCGACGATTCAGGACCATGGAGATGATTCA 663

QY 1044 TGAAGAGATGTGGCAGCA 1062
Db 664 TGAAGAGATGTGGCAGCA 682

RESULT 9
BM476961 1046 bp mRNA linear 05-FEB-2002
LOCUS
DEFINITION AGSCOURT 6481569 NIH_MGC_71 Homo sapiens cDNA IMAGE:5555523
5', mRNA sequence.
ACCESSION BM476961
VERSION BM476961.1 GI:18526003
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1046)
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1CML2275 row: n column: 04
High quality sequence stop: 622.

FEATURES
source
location/Qualifiers

1..1046
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5555523"

/clone_lib="NIH_MGC_71"
/tissue_type="Telomysarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.1 kb."

BASE COUNT 318 a 217 c 227 g 282 t 2 others
ORIGIN

Query Match 18.6%; Score 626; DB 13; Length 1046;
Best Local Similarity 97.7%; Pred. No. 1.2e-123;

Matches 635; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 172 TCGATGGTATTTCCACAGACAGAAATTCGCTATCTGCTTCAGAGCAG 231

Db 1 TCGATGGTATTTCCACAGACAGAAATTCGCTATCTGCTTCAGAGCAG 60

QY 232 GTGAAAATGTATCCAGGTGAGGCTGTGGAGTACCTGCTTGTGAGAG 291

Db 61 GTGAAAATGTATCCAGGTGAGGCTGTGGAGTACCTGCTTGTGAGAG 120

QY 292 GTGAAGGACAGATTCAGAGCAGTGGCACCCTCGGGACATGCAGCAATTCAC 351

Db 121 GTGAAGGACAGATTCAGAGCAGTGGCACCCTCGGGACATGCAGCAATTCAC 180

QY 352 CTGTGAGACCTTGGAGAGGAGTGTGGACCTTGTGAGCTGGGAAATCGTGA 411

Db 181 CTGTGAGACCTTGGAGAGGAGTGTGGACCTTGTGAGCTGGGAAATCGTGA 240

QY 412 GGCCTCGGAGAACCGGCGCCCTGTGGCGCGCGGTACATGAACCTGAGCTCA 471

Db 241 GGCCTCGGAGAACCGGCGCCCTGTGGCGCGCGGTACATGAACCTGAGCTCA 300

QY 472 TTGCCCTTCATCGTTTGAAGAGCTCATGATGATATCTCAACTGCTGAGCT 531

Db 472 TTGCCCTTCATCGTTTGAAGAGCTCATGATGATATCTCAACTGCTGAGCT 531

Db 301 TTGGCTCTCCATGCTTTTGAGACGCTCATGATGATATCTCCAACTGTTGAACTCTT 360
QY 532 CAGCCCACTCTGTGACAAAGCTTGTAGAGAGCTTGTGATAGTATGAGAGAG 591
Db 361 CAGCCCACTCTGTGACAAAGCTTGTAGAGAGCTTGTGATAGTATGAGAGAG 420
QY 592 GAACCTGTGCAATGAGACGAAACCGGATGTGTGAGCAAAACAAATGGAATGA 651
Db 421 GAACCTGTGCAATGAGACGAAACCGGATGTGTGAGCAAAACAAATGGAATGA 480
QY 652 TCAGGTGTAGAGAGCTACTAAAAAGATTTGTGCAAGAAACCTGTGTCTGCAATT 711
Db 481 TCAGGTGTAGAGAGCTACTAAAAAGATTTGTGCAAGAAACCTGTGTCTGCAATT 540
QY 712 CTGAATGTCTTGTGCAACAGCAAAATGAATCTGTGCAAGATTAAGCTGTGAT 771
Db 541 CTGAATGTCTTGTGCAACAGCAAAATGAATCTGTGCAAGATTAAGCTGTGAT 600
QY 772 TGCTCAAGAAAGCAATGCAAGATTTGAGAAATTTATGCAAGATTTGAGTGTG 821
Db 601 TGCTCAAGAAAGCAATGCAAGATTTGAGAAATTTATGCAAGATTTGAGTGTG 650

RESULT 10
BE882040 1041 bp mRNA lin. 1 EST 20-OCT-2000
LOCUS 601505326P2 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3906774 5',
DEFINITION mRNA sequence.
ACCESSION BE882040
VERSION BE882040.1 GI:10330816
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 1041)
AUTHORS NIH-MGC http://mgs.scl.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LAM9716 row: d column: 07
High quality sequence stop: 586.

FEATURES
Location/Qualifiers
1..1041

ORGANISM="Homo sapiens"
/db_xref="taxon:9606"
/clone IMAGE:3906774
/clone_lib="NIH_MGC_71"
/tissue_type="Telomysarcoma"
/lab_host="DH10B (phage-res. stand)"
/note="Organ: uterus; Vector: pCMV-Sport6; Site: 1; Note:
Site 2: Salt: Cloned unidirectionally. Primer: Oligo dt.
Average insert size 2.1 kb."
BASE COUNT 385 a 258 c 223 g 175 t
ORIGIN

Query Match 18.5% Score 622.8; DB 12; Length 1041;
Best Local Similarity 89.1%; Pred. No. 5.9e-123;
Matches 718; Conservative 0; Mismatches 82; Indels 6; Gaps 4;
QY 1923 GGAAGAAAAAGTGTCAAAAAAGAAATGCAAAAGACGTGTTGTGCAATGATTGAG 1982
Db 1 GGAAGAAAAAGTGTCAAAAAAGAAATGCAAAAGACGTGTTGTGCAATGATTGAG 60
QY 1983 GAAGTCAATGAGGCGCTTCAAAATTAA-TGACACAATTGCAATGATAGATGATATCTC 2041

Db 61 GAAGTACAAAGGAGGCGCTTCAAAATTAAAGACACAAATGGAATGATGCTATATCTC 120
QY 2042 ATCTTGAAACTTCTATATATGAGAAAGAAATGAAAGCTTGTGCAATGATGATA 2101
Db 121 ATCTTGAAACTTCTATATATGAGAAAGAAATGAAAGCTTGTGCAATGATGATA 180
QY 2102 GTGATGAGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2161
Db 181 GTGATGAGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
QY 2162 AACCTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2221
Db 241 AACCTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
QY 2222 AAATGTGAAAGGCTGTGTAAGAAACCGAATATGAAATGAAAGCTGACCAATTTA 2281
Db 301 AAATGTGAAAGGCTGTGTAAGAAACCGAATATGAAATGAAAGCTGACCAATTTA 360
QY 2282 GAAATACCATTAATGAGCAATATGATGATGATGATGATGATGATGATGATGAT 2341
Db 361 GAAATACCATTAATGAGCAATATGATGATGATGATGATGATGATGATGATGAT 420
QY 2342 CAAAAACAGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2401
Db 421 CAAAAACAGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
QY 2402 CTGAAGTAGAGTCAAAAGCCCAATCTGATGATGATGATGATGATGATGATGAT 2461
Db 481 CTGAAGTAGAGTCAAAAGCCCAATCTGATGATGATGATGATGATGATGATGAT 540
QY 2462 CCATGACACAGATATGACAAAAAGAGTATGATGATGATGATGATGATGATGAT 2520
Db 541 CCATGACACAGATATGACAAAAAGAGTATGATGATGATGATGATGATGATGAT 600
QY 2521 CTGCTTATGCTTACGACAGATGATGATGATGATGATGATGATGATGATGAT 2580
Db 601 CTGCTTATGCTTACGACAGATGATGATGATGATGATGATGATGATGATGAT 660
QY 2581 ATCCGTTATGCTTACGACAGATGATGATGATGATGATGATGATGATGATGAT 2640
Db 661 ATCCGTTATGCTTACGACAGATGATGATGATGATGATGATGATGATGATGAT 717
QY 2641 CCTGATGAGAGCACTAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2699
Db 718 TACTGAAAGAACCAAGACGCGCTGCGCAAAAGGCTGACAGCAGCAACCAAGCCA 777
QY 2700 GACAGTAAATGATTTCCGAGAGAGA 2725
Db 778 CACAGCCAGACATCCAGAAACA 803

RESULT 11
LOCUS AK018602
DEFINITION Mus musculus adult male cecum cDNA, RIKEN full-length enriched
library, clone:9130009C22; homolog to MELANOMA DIFFERENTIATION
ASSOCIATED PROTEIN-5, full insert sequence.
ACCESSION AK018602
VERSION AK018602.1 GI:12858393
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain: C57BL/6J) adult male cecum cDNA to mRNA,
clone: 9130009C22.
ORGANISM Mus musculus

REFERENCE 1
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 16, 2003, 09:54:24 ; Search time 114.8 seconds

(without alignment) updates/sec
9052.345 Million

Title: US-09-515-363C-1

Perfect score: 3365

Sequence: 1 gcgcgcgcgcctgagagccc.....aaatattgttttaccctg 3365

Scoring table:

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 88211

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued_Patents_NA:
1: /cgn2_6/ptodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/lna/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/lna/PCITS.COMB.seq:*
6: /cgn2_6/ptodata/1/lna/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	285.4	8.5	301	1	US-08-143-576-7 Sequence 7, Appl
2	124.6	3.7	193	1	US-08-143-576-6 Sequence 6, Appl
3	69.4	2.1	7218	1	US-08-232-463-14 Sequence 14, Appl
4	62	1.8	5852	4	US-09-853-768-10 Sequence 10, Appl
5	62	1.8	7037	4	US-09-853-768-3 Sequence 3, Appl
6	54.6	1.6	371	4	US-09-222-578-46 Sequence 46, Appl
7	49.8	1.5	8920	2	US-08-446-855A-1 Sequence 1, Appl
8	49.8	1.5	8920	4	US-09-150-741-1 Sequence 1, Appl
9	47.8	1.4	2319	3	US-09-058-489-90 Sequence 90, Appl
10	47.8	1.4	4416	3	US-09-058-489-17 Sequence 17, Appl
11	47.8	1.4	5361	4	US-08-973-462-2 Sequence 2, Appl
12	47.8	1.4	6152	4	US-08-973-462-1 Sequence 1, Appl
13	47.2	1.4	2295	1	US-08-375-300-3 Sequence 3, Appl
14	47.2	1.4	2295	5	US-09-177-431-3 Sequence 3, Appl
15	47.2	1.4	2295	5	PCT-US95-16930-3 Sequence 3, Appl
16	47.2	1.4	4080	1	US-08-375-300-1 Sequence 1, Appl
17	47.2	1.4	4080	3	US-09-177-431-1 Sequence 1, Appl
18	47.2	1.4	4080	5	PCT-US95-16930-1 Sequence 1, Appl
19	46.8	1.4	1956	4	US-08-559-896B-1 Sequence 1, Appl
20	45.8	1.4	3408	4	US-09-058-489-14 Sequence 14, Appl
21	45.8	1.4	5322	4	US-09-058-489-13 Sequence 13, Appl
22	43.8	1.3	2055	4	US-09-134-001C-1938 Sequence 1938, Ap
23	43.8	1.3	5852	1	US-07-867-106-2 Sequence 2, Appl
24	43.4	1.3	696	4	US-09-461-697-193 Sequence 193, App
25	43.4	1.3	699	4	US-09-461-697-191 Sequence 191, App
26	43.4	1.3	717	4	US-09-461-697-189 Sequence 189, App
27	43.4	1.3	774	4	US-09-461-697-187 Sequence 187, App

ALIGNMENTS

28	43.4	1.3	819	4	US-09-461-697-185	Sequence 185, App
29	43.4	1.3	1669	4	US-09-461-697-184	Sequence 184, App
30	43.2	1.3	2518	4	US-09-433-699-3	Sequence 3, Appl
31	43	1.3	966	2	US-08-766-738-2	Sequence 2, Appl
32	43	1.3	966	4	US-09-262-610-2	Sequence 2, Appl
33	42.6	1.3	435	4	US-09-397-787-171	Sequence 171, App
34	42	1.2	2277	1	US-08-676-967-2	Sequence 2, Appl
35	42	1.2	2277	1	US-08-676-974-2	Sequence 2, Appl
36	42	1.2	2277	2	US-09-098-487-2	Sequence 2, Appl
37	42	1.2	6002	4	US-09-345-882-4	Sequence 4, Appl
38	41.2	1.2	1164	4	US-09-538-871-3	Sequence 3, Appl
39	41.2	1.2	1232	4	US-09-538-871-1	Sequence 1, Appl
40	41.2	1.2	2733	1	US-08-676-967-3	Sequence 3, Appl
41	41.2	1.2	2733	1	US-08-676-974-3	Sequence 3, Appl
42	41.2	1.2	2733	2	US-09-098-487-3	Sequence 3, Appl
43	41.2	1.2	8050	4	US-09-491-362-11	Sequence 11, Appl
44	41.2	1.2	8050	4	US-09-874-562-11	Sequence 11, Appl
45	41.2	1.2	18596	4	US-09-318-448-11	Sequence 11, Appl

RESULT 1
US-08-143-576-7
Sequence 7, Application US/08143576
Patent No. 5643761
GENERAL INFORMATION:
APPLICANT: Fisher, Paul B.
APPLICANT: Jiaog, Hongping
TITLE OF INVENTION: METHOD FOR GENERATING A SUBSTRUCTURED
TITLE OF INVENTION: CNNA LIBRARY AND USES OF THE GENERATED LIBRARY
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White, c/o Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/143,576
FILING DATE: 25-OCT-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 43563/JPW/ANC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 664-0525
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 301 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-143-576-7
Query Match 8.5%, Score 285.4; DB 1; Length 301;
Best Local Similarity 99.3%; Pred. No. 1.3e-64;
Matches 297; Conservative 0; Mismatches 1; Gaps 1;
QY 1592 ACAAACGAGTATTCCTCTCTCAGATACAGGACAAAGAGCTTACCTGCTGTTTACAG 1651

US-08-446-855A-1
Sequence 1, Application US/08446855A
Patent No. 5849573
GENERAL INFORMATION:
APPLICANT: Stewart, Thomas S
APPLICANT: Flores, Maria V
APPLICANT: O'Sullivan, William J
TITLE OF INVENTION: Nucleotide sequence encoding card ...
TITLE OF INVENTION: Phosphate synthetase II
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon & Vanderhye PC
STREET: 1100 No. 5849573th Glebe Road, 8th Floor
CITY: Arlington
STATE: Virginia
COUNTRY: USA
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,855A
FILING DATE: 06-Jul-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mitchard, Leonard C
REGISTRATION NUMBER: 29,009
REFERENCE/DOCKET NUMBER: 47-80
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-816-4000
TELEFAX: 703-816-4100
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 8920 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: genomic
US-08-446-855A-1

Query Match 1.5% Score 49.8; DB 2; Length 8920;
Best Local Similarity 51.1%; Pred. No. 0.0079;
Matches 117; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 1926 AAAAAAGCTGCAAAAAAGAAATCGCAAGAACTGTTGTGAGAAATTTGAGCA 1985
DB 6394 AATAAAGAAATATTTATGAGAGTCAAAATATATATTTTACATTTTCGACCA 6453
QY 1986 GTACATGAGCCCTTCAAAATTAATGACCAATTCGAATGATACGCGTACTCATCT 2045
DB 6454 AGCTAACATGACGATTTAAGTAATCATATTTAATTAAGTACTTAAATATCTGA 6513
QY 2046 TGAACCTTCTATATGAAGAGATTAAGAGTTTGAGTATAGAAATGATATCTGA 2105
DB 6514 TGAGTATGATGATGATATTTATTTATGATGATGATGATGATGATGATGATGAT 6573
QY 2106 TGAGGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2154
DB 6574 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6622

RESULT 8
US-09-150-741-1
Sequence 1, Application US/09150741
Patent No. 6183996
GENERAL INFORMATION:
APPLICANT: Stewart et al.
TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
Patent No. 6183996
TITLE OF INVENTION: synthetase II
FILE REFERENCE:

CURRENT APPLICATION NUMBER: US/09/150,741
CURRENT FILING DATE: 1998-09-10
EARLIER APPLICATION NUMBER: PL6380
EARLIER FILING DATE: 1992-12-16
EARLIER APPLICATION NUMBER: AU93/00617
EARLIER FILING DATE: 1993-12-02
EARLIER APPLICATION NUMBER: 08/446,855
EARLIER FILING DATE: 1995-07-06
NUMBER OF SEQ ID NOS: 15
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 1
LENGTH: 8920
TYPE: DNA
ORGANISM: Plasmodium falciparum
US-09-150-741-1

Query Match 1.5% Score 49.8; DB 4; Length 8920;
Best Local Similarity 51.1%; Pred. No. 0.0079;
Matches 117; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 1926 AAAAAAGCTGCAAAAAAGAAATCGCAAGAACTGTTGTGAGCAAAATTTGAGCA 1985
DB 6394 AATAAAGAAATATTTATGAGAGTCAAAATATATATTTTACATTTTCGACCA 6453
QY 1986 GTACATGAGCCCTTCAAAATTAATGACCAATTCGAATGATACGCGTACTCATCT 2045
DB 6454 AGCTAACATGACGATTTAAGTAATCATATTTAATTAAGTACTTAAATATCTGA 6513
QY 2046 TGAACCTTCTATATGAAGAGATTAAGAGTTTGAGTATAGAAATGATATCTGA 2105
DB 6514 TGAGTATGATGATGATATTTATTTATGATGATGATGATGATGATGATGATGAT 6573
QY 2106 TGAGGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2154
DB 6574 TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6622

RESULT 9
US-09-058-489-90
Sequence 90, Application US/09058489
Patent No. 6103886
GENERAL INFORMATION:
APPLICANT: Whitehead Institute for Biomedical Research
APPLICANT: Lahn, Bruce
APPLICANT: Page, David
TITLE OF INVENTION: Genes in the No. 6103886-Recombining Protein of
TITLE OF INVENTION: the Y Chromosome
FILE REFERENCE: WH197-08PA
CURRENT APPLICATION NUMBER: US/09/058,489
CURRENT FILING DATE: 1998-04-10
EARLIER APPLICATION NUMBER: 60/041,877
EARLIER FILING DATE: 1997-04-11
NUMBER OF SEQ ID NOS: 91
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 90
LENGTH: 2319
TYPE: DNA
ORGANISM: Human
US-09-058-489-90

Query Match 1.4% Score 47.8; DB 3; Length 2319;
Best Local Similarity 53.5%; Pred. No. 0.013;
Matches 100; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 2466 GACACAGATGACAAAGAAAGTATTAATTTGCGAGTGAATAATCAATTCGCT 2525
DB 1490 GTCAAGAGAGATGACAGAGAGGCGCTTCACCAAGTTGCTGACAGAAAGTCCAAATTC 1549
QY 2526 TATGCTACACAGTGGCAGAGAAAGTCTGATATTAAAGCAATGTAACATTTGATCG 2585
DB 1550 AGTGCTACAGCTGTGGCAGAGAGGAGTACGATTTCAAAATGATGAGACATGTTATCAA 1609
QY 2586 TTATGCTGTGTCACCAATGAATAATGATGCTGACAGCCGCTGTCAGTACACTGA 2645

STREET: 225 Franklin Street Suite 3100
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/16930
FILING DATE: 27-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/375,300
FILING DATE: 20-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, J. Peter
REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 04020/046W01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)542-5070
TELEFAX: (617)542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2295 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
PCT-US95-16930-3

Query Match 1.48; Score 47.2; DB 5; Length 2295;
Best Local Similarity 51.48; Pred. No. 0.019;
Matches 109; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 2057 ATAATGAGAGAAAGATTAAGAGTTTGCAAGTATAGAGATGATGATGAGGCGTG 2116
DB 1718 ATGACGATGACGATGACGATGATGATGATGATGAGAGAGGCGATGATGATGATG 1777
QY 2117 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2176
DB 1778 AAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1837
QY 2177 ATGAAACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2236
DB 1838 AGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1897
QY 2237 TGGCTGAAGCCAGCAATATGAAATGAAAG 2268
DB 1898 ACGAGAGAAACTTAAGAGATGAGAGAAAGG 1929

Search completed: May 16, 2003, 13:24:45
Job time : 294 secs

PN MO200164707-A1.
 XX
 PD 07-SEP-2001.
 XX
 PF 28-FEB-2001; 2001WO-US06960.
 XX
 PR 29-FEB-2000; 2000US-0515363.
 XX
 PA (UYCO) UNIV COLUMBIA NEW YORK.
 XX
 PI Fisher PB, Kang D, GopalKrishnan RV.
 XX
 DR WPI: 2001-565494/63.
 DR P-PSDB: AAE10155.
 XX
 PT Nucleic acid sequences encoding a Melanoma Differentiation Associated
 PT Gene useful for cancer cell growth suppression, apoptosis and
 PT anti-viral activity -
 XX
 PS Claim 1; Page 16-18; 152pp; English.
 XX
 CC The present invention relates to an isolated nucleic acid encoding a
 CC melanoma differentiation associated gene-5 (Mda-5) polypeptide. Mda-5
 CC contains a caspase recruitment domain (CARD) and a RNA helicase motif.
 CC Mda-5 is a novel interferon (IFN) inducible gene with structural
 CC similarities to RNA helicases and CARD motif containing proteins. Mda-5
 CC is induced during terminal differentiation in human melanoma cells
 CC treated with the combination of recombinant fibroblast IFN and the
 CC antileukemic compound mezerein (MEZ). Mda-5 is useful for identifying
 CC compounds that may induce its expression. Mda-5 is useful for treating
 CC cancer e.g., melanoma, neuroblastoma, astrocytoma, glioblastoma
 CC multiforme, cervical cancer, breast cancer, colon cancer, prostate
 CC cancer, osteosarcoma, chondrosarcoma and especially where the cancer is
 CC a cancer of the central nervous system and apoptosis. The Mda-5 promoter
 CC exhibits melanocyte tissue specificity and minimises systemic toxicity.
 CC The present sequence is human Mda-5 cDNA.
 XX
 SO Sequence 3365 BP; 1138 A; 644 C; 753 G; 830 T; 0 other.

Query Match 100.0%; Score 3365; DB 22; Len 1365;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGGGGGGGGCTGAGAGCCCTGAGCAACCTGATATGTCAGACACACACGATAGC 60
 DB 1 GGGGGGGGGCTGAGAGCCCTGAGCAACCTGATATGTCAGACACACACGATAGC 60
 QY 61 CCGTCTTCTTAAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 120
 DB 61 CCGTCTTCTTAAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 120
 QY 121 CAGCACCATCTGCTGGGAGAACCTCTCCCTTCTGTGAGAAAGAAAGATGGAA 180
 DB 121 CAGCACCATCTGCTGGGAGAACCTCTCCCTTCTGTGAGAAAGAAAGATGGAA 180
 QY 181 TATTCACAGAGAGAGAAATTCGCGTATCTCATCTGCTGTTAGAGGCCAAGTGA 240
 DB 181 TATTCACAGAGAGAGAAATTCGCGTATCTCATCTGCTGTTAGAGGCCAAGTGA 240
 QY 241 TACATCAGAGTGGAGCTGCTGAGTACCTGACCTTCTGCTGAGCAACATGAGAG 300
 DB 241 TACATCAGAGTGGAGCTGCTGAGTACCTGACCTTCTGCTGAGCAACATGAGAG 300
 QY 301 CAGATTCAGAGGACAGTGGCCACCTCCGGGAACATGACAGGCAATGAAATGCTGA 360
 DB 301 CAGATTCAGAGGACAGTGGCCACCTCCGGGAACATGACAGGCAATGAAATGCTGA 360
 QY 361 ACCTTGAGAGAGAGAGTCTGGGACCTTGGTGGAGTCCGGAATGGTGAAGCCCTCG 420
 DB 361 ACCTTGAGAGAGAGAGTCTGGGACCTTGGTGGAGTCCGGAATGGTGAAGCCCTCG 420
 QY 421 AGAAGCGGAGCGCTCTGGCGCCGCTCATGAAACCCGAGCTACGAGTTCGCGCTCT 480
 DB 421 AGAAGCGGAGCGCTCTGGCGCCGCTCATGAAACCCGAGCTACGAGTTCGCGCTCT 480

DB 421 AGAAGCGGAGCGCTCTGGCGCCGCTCATGAAACCCGAGCTACGAGTTCGCGCTCT 480
 QY 481 CCATGGTTTGAAGAGCGCTCATGAAATATCTCAACTGCTGAAACCTCTTCAGCCACT 540
 DB 481 CCATGGTTTGAAGAGCGCTCATGAAATATCTCAACTGCTGAAACCTCTTCAGCCACT 540
 QY 541 CTGCTGGAGACGCTCTTACTAGACAGCTCTTGCAATAGTGCAATGAGAGAGACTGTTG 600
 DB 541 CTGCTGGAGACGCTCTTACTAGACAGCTCTTGCAATAGTGCAATGAGAGAGACTGTTG 600
 QY 601 ACAATTGAAAGACAGAAACCGGATGCTGTCAGAAACCAATGAAATGAAATCAGTCTA 660
 DB 601 ACAATTGAAAGACAGAAACCGGATGCTGTCAGAAACCAATGAAATGAAATCAGTCTA 660
 QY 661 AGAGAGCTACTAAAAAGATTGTGCAGAAAGAAACCTGCTCTGCAATTTCTGAAATGT 720
 DB 661 AGAGAGCTACTAAAAAGATTGTGCAGAAAGAAACCTGCTCTGCAATTTCTGAAATGT 720
 QY 721 CTTGCTCAAAACAGAAACATGAACTGCTCCAAAGATTACAGGCTCTGATTCACAGAA 780
 DB 721 CTTGCTCAAAACAGAAACATGAACTGCTCCAAAGATTACAGGCTCTGATTCACAGAA 780
 QY 781 ACCAATGAGAGATGAGATTTATCACAAGTTGATGCTCTCAAGTGAAGAGCAACTT 840
 DB 781 ACCAATGAGAGATGAGATTTATCACAAGTTGATGCTCTCAAGTGAAGAGCAACTT 840
 QY 841 CTTTCAACACAGTTCAGCAAAATCTGCAGAAAGAGTCTGGGATGAGAAATATCTCA 900
 DB 841 CTTTCAACACAGTTCAGCAAAATCTGCAGAAAGAGTCTGGGATGAGAAATATCTCA 900
 QY 901 TCGAATCATCTTTTGCAGATTTCTCTAGTTTCAAGTATGAGACACAAAGTTTGGCAGAA 960
 DB 901 TCGAATCATCTTTTGCAGATTTCTCTAGTTTCAAGTATGAGACACAAAGTTTGGCAGAA 960
 QY 961 GGAATGTCAGTGTGTTGATGAAAGTCTGAGCAATTAACAGCAATGAGGAGTATTC 1020
 DB 961 GGAATGTCAGTGTGTTGATGAAAGTCTGAGCAATTAACAGCAATGAGGAGTATTC 1020
 QY 1021 GGCACCATGAGAGATGATTCAGATGAAAGATGTGGCAGCAAGCATCCCGAGGCA 1080
 DB 1021 GGCACCATGAGAGATGATTCAGATGAAAGATGTGGCAGCAAGCATCCCGAGGCA 1080
 QY 1081 GAATTCACAGTCTAGGCTTTACCAATGGAATGTGGCCAGCCAGCTTGGAAAGGAGAAAT 1140
 DB 1081 GAATTCACAGTCTAGGCTTTACCAATGGAATGTGGCCAGCCAGCTTGGAAAGGAGAAAT 1140
 QY 1141 ATCATCATCTGCTCCCTACAGAGAGTGAAGAAACCAAGAGTGGCTTTACATTTGCCAAG 1200
 DB 1141 ATCATCATCTGCTCCCTACAGAGAGTGAAGAAACCAAGAGTGGCTTTACATTTGCCAAG 1200
 QY 1201 GATCACTTAGACAGAGAAAGAGCATGTGAGCCTGGAAGAAATTAATGTTCTGTCAAT 1260
 DB 1201 GATCACTTAGACAGAGAAAGAGCATGTGAGCCTGGAAGAAATTAATGTTCTGTCAAT 1260
 QY 1261 AAGTACTGCTAGTGAAGACAGCTCTTCGCAAGGAGTTCACCAATTTTGAAGAAATGG 1320
 DB 1261 AAGTACTGCTAGTGAAGACAGCTCTTCGCAAGGAGTTCACCAATTTTGAAGAAATGG 1320
 QY 1321 TATCGTGTATTTGATTAAGTGAAGTGAATCCCAACGAAATATATATTTCCAGAAAGTGTG 1380
 DB 1321 TATCGTGTATTTGATTAAGTGAAGTGAATCCCAACGAAATATATATTTCCAGAAAGTGTG 1380
 QY 1381 AAGTCTGTGATTAATTAATAGTACAGCTCAAAATCTTGAAGAAATCCCTCTTAATCTTG 1440
 DB 1381 AAGTCTGTGATTAATTAATAGTACAGCTCAAAATCTTGAAGAAATCCCTCTTAATCTTG 1440
 QY 1441 GAAATGAGAGAGAGTCTGCTGCTTCAATTTGCAAGCTTTTCCCATATATATGATGAGAA 1500
 DB 1441 GAAATGAGAGAGAGTCTGCTGCTTCAATTTGCAAGCTTTTCCCATATATATGATGAGAA 1500
 QY 1501 TGTCAATCAGCAACAGAGAGAGTGTATATTAACATCATGAGGAGATATTTGATTCAG 1560
 DB 1501 TGTCAATCAGCAACAGAGAGAGTGTATATTAACATCATGAGGAGATATTTGATTCAG 1560

FT	/*tag- a	
ET	/product- "Human RH16"	
XX	WO200185955-A1.	
XX		
XX	15-NOV-2001.	
XX		
XX	11-MAY-2001; 2001WO-FR01441.	
XX		
XX	11-MAY-2000; 2000FR-0006030.	
XX		
XX	(ISTA-) ISTAC.	
XX	(INSP) INST PASTEUR LILLE.	
XX		
PI	Bahr G, Cocude C, Capron A;	
XX		
DR	WPI; 2002-082898/11.	
DR	P-PSDB; AAM47798.	
XX		
PT	New polypeptide, useful for treating and diagnosing cancer or	
PT	inflammation, and drug screening, comprises a human polynucleotide	
PT	homologous to RNA helicase	
XX		
PS	Claim 7; Page 85-89; 114pp; French.	
XX		
CC	The present sequence is the coding sequence for human RH16. RH16 is a	
CC	10kDa protein and has homology to RNA helicases (DEX1 box). RH16 and	
CC	its coding sequence are useful for treating cancer; auto or chronic	
CC	infections (especially by HIV or hepatitis B or C); inherited genetic	
CC	diseases; (auto)immune diseases (particularly rheumatism, arthritis,	
CC	arteriosclerosis, osteoporosis and diabetes, but many others listed) and	
CC	to prevent graft rejection. RH16 and its coding sequence are also useful	
CC	for inducing, or increasing, the immune response to a vaccine.	
XX		
SQ	Sequence 3372 BP; 1157 A; 637 C; 748 G; 830 T; 0 other;	
Query Match	99.2%; Score 3338.8; DB 24; Length 3372;	
Best Local Similarity	99.8%; Pred. No. 0;	
Matches 3343; Conservative	0; Mismatches 7; Indels 0; Gaps 0	
0Y	15 GAGCCCTGTGACAACTCGTCATTGTGCAGGCAGAGCGGTATACCCCTCTTCTCTAG 74	
Db	1 GGGCCCTGTGACAACTCGTCATTGTGCAGGCAGAGCGGTATACCCCTCTTCTCTAG 60	
0Y	75 TGGGACGGGAGAGCGGCAGCAGCATTTTCACCTGTCCCTAGACACAGAGCATCTGCT 134	
Db	61 TGGGACGGGAGAGCGGCAGCAGCATTTTCACCTGTCCCTAGACACAGAGCATCTGCT 120	
0Y	135 TGGGAGAACCCCTCTCCCTCTGTGAGAAAGAAAGATGTGAAATGGGTATCTGACAGCA 194	
Db	121 TGGGAGAACCCCTCTCCCTCTGTGAGAAAGAAAGATGTGAAATGGGTATCTGACAGCA 180	
0Y	195 GAATTTCCGCTATTCATCTCGTGGTTCAGGGCCAGGGGTGAATATGTATCTCAGGTGA 254	
Db	181 GAATTTCCGCTATTCATCTCGTGGTTCAGGGCCAGGGGTGAATATGTATCTCAGGTGA 240	
0Y	235 GCGTGTGTGACACTACCTGCTTCTGCTGCTTCCAGAGGTGAAGGAGCAGATTCAGAGGAC 314	
Db	241 GCGTGTGTGACACTACCTGCTTCTGCTGCTTCCAGAGGTGAAGGAGCAGATTCAGAGGAC 300	
0Y	315 AGTCCCACTCTCGGGAAATATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 374	
Db	301 AGTCCCACTCTCGGGAAATATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360	
0Y	375 AGTCTGGACACTTGGTGGACTCGGGAATTTGGTGGAGGCGCCCGGAGAGATTTGGCAGGCC 434	
Db	361 AGTCTGGACACTTGGTGGACTCGGGAATTTGGTGGAGGCGCCCGGAGAGATTTGGCAGGCC 420	
0Y	435 TCTGGCGCGCCCTCATATGAAACCTGTAGCTCAGGACTTGGCTCTCCATCTTTTGAAGA 494	
Db	421 TCTGGCGCGCCCTCATATGAAACCTGTAGCTCAGGACTTGGCTCTCCATCTTTTGAAGA 480	
0Y	495 GCGTCATGATGAAATCTTCCAACTGCTGAACCTGCTTACGCCCACTCTGTTAGCAACCT 554	

Db	481	CGCTCATSATAAATATCTCCACACTGCTGAACCTCTTCAGCCCACTCTGGTGGAAACACT	540
QY	555	TCCTAGTTGAGACAGCTCTTGGATTAATGTCATGAGGAGGAGAACTGTTGACATTTGAAGACAG	614
Db	541	TCTAGTTAGAACAGCTCTTGGATTAATGTCATGAGGAGGAGAACTGTTGACATTTGAAGACAG	600
QY	615	AAACCGCATTCCTCTCTGACAGAAAACAATGSAATGAAATCAGTGTGTAAGACAGCTACTAA	674
Db	601	AAACCGCATTCCTCTCTGACAGAAAACAATGSAATGAAATCAGTGTGTAAGACAGCTACTAA	660
QY	675	AAGGATTTGTGAGAAAGAAAAGCTGGTCTCTGCAATTTGGAATGTTCTCTCTAAACAGS	734
Db	661	AAGGATTTGTGAGAAAGAAAAGCTGGTCTCTGCAATTTGGAATGTTCTCTCTAAACAGS	720
QY	735	AAACATGTAACCTTGTCCAGAGAGTTAACAGGCTCTGATTTGCTGAGAAAGCAATGACAGAT	794
Db	721	AAACATGTAACCTTGTCCAGAGAGTTAACAGGCTCTGATTTGCTGAGAAAGCAATGACAGAT	780
QY	795	TGAGCAATTTATCACAGTGTATGCTCTCAATGAGAAAGCAACTTTCTTCAACACAGT	854
Db	781	TGAGCAATTTATCACAGTGTATGCTCTCAATGAGAAAGCAACTTTCTTCAACACAGT	840
QY	855	TCAGCAAAATGTGAGAAAGAGGCTGGGCAATGAGAAATACCTATCATGAAATCATCTT	914
Db	841	TCAGCAAAATGTGAGAAAGAGGCTGGGCAATGAGAAATACCTATCATGAAATCATCTT	900
QY	915	TGCAGATTCCTCTAGTTTCACAAATCAGACAACTTTGGGAGAAAGAGTCAAGCTG	974
Db	901	TGCAGATTCCTCTAGTTTCACAAATCAGACAACTTTGGGAGAAAGAGTCAAGCTG	960
QY	975	CTTAGATGAAAGCTTGTGGACATAACAGCAACATGAGCAAGTATTCAGGACCTACGGAAAG	1034
Db	961	CTTAGATGAAAGCTTGTGGACATAACAGCAACATGAGCAAGTATTCAGGACCTACGGAAAG	1020
QY	1035	TGATTCAATGAATAAGAAATGTGGCAGCAAGAGCATCCCGGAGCCAGAACTTCACCTAG	1094
Db	1021	TGATTCAATGAATAAGAAATGTGGCAGCAAGAGCATCCCGGAGCCAGAACTTCACCTAG	1080
QY	1095	GCCTTACCAAAATGGAAGTTGCCAGCCAGCCTTGGAAAGGAGAAATATCATCATCTGCT	1154
Db	1081	GCCTTACCAAAATGGAAGTTGCCAGCCAGCCTTGGAAAGGAGAAATATCATCATCTGCT	1140
QY	1155	CCCTTACAGGAGGTGGAAAAAACACAGAGTGGCTGTTTACATTTGCCAAGATTCACCTAGACA	1214
Db	1141	CCCTTACAGGAGGTGGAAAAAACACAGAGTGGCTGTTTACATTTGCCAAGATTCACCTAGACA	1200
QY	1215	GAAAGAAAAACATCTGAGGCTGSAAGATTAAGTCTTGTCATATAACGTCTCTCTAGT	1274
Db	1201	GAAAGAAAAACATCTGAGGCTGSAAGATTAAGTCTTGTCATATAACGTCTCTCTAGT	1260
QY	1275	TGAACAGCTCTTCCGCAAGAGATTGCCAACCATTTTGAAGAAATGGTATGATATGG	1334
Db	1261	TGAACAGCTCTTCCGCAAGAGATTGCCAACCATTTTGAAGAAATGGTATGATATGG	1320
QY	1335	ATTAAAGTGTGATATCCCACTGAAAAATATCATTTCCAGAAAGTTGTCAAGTCTGTGATAT	1394
Db	1321	ATTAAAGTGTGATATCCCACTGAAAAATATCATTTCCAGAAAGTTGTCAAGTCTGTGATAT	1380
QY	1395	TATTATCAGTACACTCAAAATCTGTGAAACTCCCTTAAACTTGGAAAAAGGAGAGA	1454
Db	1381	TATTATCAGTACACTCAAAATCTGTGAAACTCCCTTAAACTTGGAAAAAGGAGAGA	1440
QY	1455	TGCTGGTGTTCATATGTCAGAGCTTTTCCCATATTCATGATGAAATGTCAATCAGCTAA	1514
Db	1441	TGCTGGTGTTCATATGTCAGAGCTTTTCCCATATTCATGATGAAATGTCAATCAGCTAA	1500
QY	1515	CAAAAGAACAGTGTATATATACATCATGAGGCAATTAATTTGATGACAAAGCTGAAAAACAA	1574
Db	1501	CAAAAGAACAGTGTATATATACATCATGAGGCAATTAATTTGATGACAAAGCTGAAAAACAA	1560
QY	1575	TAGACTCAAGAAAGAAAACAACAGTGAATCCCTTCTCAGATATCTGGCAATTAAGC	1634

Db 1561 TAGACTCAGAGAAAGAAACCAACAGTATGTCCTCTCAGACTCTGTAAGTACAGC 1620
 QY 1635 TTGACCTGGTGTGGAGGGGCGACGAGCAAGCAAGCTGAGAACAGCCTTTAAAGCT 1694
 Db 1621 TTACCTGGTGTGGAGGGGCGACGAGCAAGCAAGCTGAGAACAGCCTTTAAAGCT 1680
 QY 1695 ATGTGGCAATCTTGATGACTTACTATTTAAAGTGTAAAGAAAGCTTATGCACTGAA 1754
 Db 1681 ATGTGGCAATCTTGATGACTTACTATTTAAAGTGTAAAGAAAGCTTATGCACTGAA 1740
 QY 1755 AAACCAATACAGAGAGCAATGCAAGAGTTGGCAATGCAAGTACAGTCC 1814
 Db 1741 AAACCAATACAGAGAGCAATGCAAGAGTTGGCAATGCAAGTACAGTCC 1800
 QY 1815 ATTTAAGAGAACTCTAGCAATATGCAAGAGTTCAACTATTTGTAATGAGTCC 1874
 Db 1801 ATTTAAGAGAACTCTAGCAATATGCAAGAGTTCAACTATTTGTAATGAGTCC 1860
 QY 1875 AATGTGAGATTTTGAACCTCAGCCTATGCAACATGGCCATTCAAATG 1934
 Db 1861 AATGTGAGATTTTGAACCTCAGCCTATGCAACATGGCCATTCAAATG 1920
 QY 1935 TGCAGAAAAAGAAATCGCAAGACGCTTTTGTGCAAGCAATTTGAGTACAGTGA 1994
 Db 1921 TGCAGAAAAAGAAATCGCAAGACGCTTTTGTGCAAGCAATTTGAGTACAGTGA 1980
 QY 1995 GGGCCCTACAAATTAAGCAATTCGAATGATAGATGCTATACCACTTGTGAACCTT 2054
 Db 1981 GGGCCCTACAAATTAAGCAATTCGAATGATAGATGCTATACCACTTGTGAACCTT 2040
 QY 2055 CTATTAATGAGAGAAAGATTAAGAGTTTGCAGTCATGAGATGATAGTATGAGGGTGG 2114
 Db 2041 CTATTAATGAGAGAAAGATTAAGAGTTTGCAGTCATGAGATGATAGTATGAGGGTGG 2100
 QY 2115 TGATGATGACTATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2174
 Db 2101 TGATGATGACTATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2160
 QY 2175 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2234
 Db 2161 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2220
 QY 2235 GCTGGCTGAGAAAGCCAGCAATATGAAAGTGCAGCAATTAAGATACCAATAT 2294
 Db 2221 GCTGGCTGAGAAAGCCAGCAATATGAAAGTGCAGCAATTAAGATACCAATAT 2280
 QY 2295 GAGCAATATATAGAGTGAAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGAT 2354
 Db 2281 GAGCAATATATAGAGTGAAGATGAGAGATGAGAGATGAGAGATGAGAGATGAGAGAT 2340
 QY 2355 GAGTGCATATGCGCTTCCAGTGGATGATGATGATGATGATGATGATGATGATGATGAT 2414
 Db 2341 GAGTGCATATGCGCTTCCAGTGGATGATGATGATGATGATGATGATGATGATGATGAT 2400
 QY 2415 CAAAGCCAGCAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2474
 Db 2401 CAAAGCCAGCAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2380
 QY 2475 TGAACAAAAAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2534
 Db 2461 TGAACAAAAAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2520
 QY 2535 CACAGTGGCAGAAAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2594
 Db 2521 CACAGTGGCAGAAAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2580
 QY 2595 CAGTACCAATGAAATAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2654
 Db 2581 CAGTACCAATGAAATAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2640
 QY 2655 CTAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2714
 Db 2641 CTAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2700

QY 2715 CCGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2774
 Db 2701 CCGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2760
 QY 2775 TGCCTATAGATTTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2834
 Db 2761 TGCCTATAGATTTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2820
 QY 2835 GAGAAATATGCGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2894
 Db 2821 GAGAAATATGCGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2880
 QY 2895 CTGCACTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2954
 Db 2881 CTGCACTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2940
 QY 2955 CAAATATGACCCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3014
 Db 2941 CAAATATGACCCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3000
 QY 3015 GAAGTGTGCGCACTATCAATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 3074
 Db 3001 GAAGTGTGCGCACTATCAATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 3060
 QY 3075 AACCAATGATGCGCAAGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 3134
 Db 3061 AACCAATGATGCGCAAGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 3120
 QY 3135 GGTTCCTCAAAATTAATTCACAAAGAAACATACAAAGAGTGGATTAATGATGATGATGAT 3194
 Db 3121 GGTTCCTCAAAATTAATTCACAAAGAAACATACAAAGAGTGGATTAATGATGATGATGATGAT 3180
 QY 3195 ATTTCCCAATCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3254
 Db 3181 ATTTCCCAATCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3240
 QY 3255 TGAAGATGCTTTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3314
 Db 3241 TGAAGATGCTTTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3300
 QY 3315 ATTATGCTACAGAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3364
 Db 3301 ATTATGCTACAGAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3250
 RESULT 3
 AAD11170
 ID AAD11170 standard; DNA: 3131 BP.
 AC XX
 NC AAD11170;
 DT 29-NOV-2001 (first entry)
 DE XX
 DE Human melanoma differentiation associated-5 protein-related DNA.
 XX XX
 KW Human melanoma differentiation associated gene; Mda-5; interferon; IFN;
 KW RNA helicase motif; caspase recruitment domain; CARD; thymic; melanoma;
 KW neuroblastoma; astrocytoma; glioblastoma; multiforme; cervical;
 KW breast; colon; prostate; osteosarcoma; chondrosarcoma; systemic toxicity;
 KW central nervous system; cystostatic; apoptosis; ds.
 XX XX
 OS Homo sapiens.
 OS XX
 PN WO200164707-A1.
 PN XX
 PD 07-SEP-2001.
 PD XX
 PF 28-FEB-2001; 2001MO-US06960.
 PF XX
 PR 29-FEB-2000; 2000US-051363.
 PR XX
 PA (UYCO) UNIV COLUMBIA NEM YORK.

XX Fisher PB, Kang D, GopalKrishnan RV;
 XX WPI: 2001-565494/63.
 XX
 PT Nucleic acid sequences encoding a Melanoma Differentiation Associated
 PT Gene useful for cancer cell growth suppression, apoptosis and
 PT anti-viral activity -
 XX
 XX Disclosure: Page 134-148; 152pp; English.
 XX
 CC The present invention relates to an isolated nucleic acid encoding a
 CC melanoma differentiation associated gene-5 (Mda-5) polypeptide. Mda-5
 CC contains a caspase recruitment domain (CARD) and a RNA helicase motif.
 CC Mda-5 is a novel Interferon (IFN) inducible gene with structural
 CC similarities to RNA helicases and CARD motif containing proteins. Mda-5
 CC is induced during terminal differentiation in human melanoma cells
 CC treated with the combination of recombinant fibroblast IFN and the
 CC anti-leukemic compound mezerein (MEZ). Mda-5 is useful for identifying
 CC compounds that may induce its expression. Mda-5 is useful for treating
 CC cancer e.g., melanoma, neuroblastoma, astrocytoma, glioblastoma
 CC multiforme, cervical cancer, breast cancer, colon cancer, prostate
 CC cancer, osteosarcoma, chondrosarcoma and especially where the cancer is
 CC a cancer of the central nervous system and apoptosis. The Mda-5 promoter
 CC exhibits melanocyte tissue specificity and minimises systemic toxicity.
 CC The present sequence is human Mda-5 protein-related DNA.
 CC Note: The present sequence is designated SEQ ID NO:2 in the sequence
 CC listing, but does not correspond to the sequence designated SEQ ID NO:2
 CC in the main body of the specification (AAE10155). The present sequence is
 CC not further referred to in the specification, and has been represented
 CC in a protein format in the sequence listing.
 XX
 SO Sequence 3131 BP; 1075 A; 583 C; 699 G; 774 T; 0 other;
 Query Match 92.2%; Score 3104; DB 22; Length 3131;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 3129; Conservative 0; Mismatches 0; Indels 5; Gaps 2;

QY 709 TTCTGAATGTTCTTCTGTCAAACAGAGAAACAATGACTTGTCCAGAGCTTAACAGCTCT 768
 DB 541 TTCTGAATGTTCTTCTGTCAAACAGAGAAACAATGACTTGTCCAGAGCTTAACAGCTCT 600
 QY 769 GATTGCTCAGAAAGCAATGCAAGATGTAAGAAATTTATACAAAGTGTGCTTCAAGT 828
 DB 601 GATTGCTCAGAAAGCAATGCAAGATGTAAGAAATTTATACAAAGTGTGCTTCAAGT 660
 QY 829 GAAGAGCAACTCTTTTCAACGACAGTTCAGCAATCTGAGAAAGAGGTGTGGGGCATG 888
 DB 661 GAAGAGCAACTCTTTTCAACGACAGTTCAGCAATCTGAGAAAGAGGTGTGGGGCATG 720
 QY 889 GAGATTAAGTATCAGAAATCATCTTTTGCAGATTTCTTGTAGTTTCAGAAATCAGACACA 948
 DB 721 GAGATTAAGTATCAGAAATCATCTTTTGCAGATTTCTTGTAGTTTCAGAAATCAGACACA 780
 QY 949 ACTTTGGCAGAAAGCAAGTGTGAGTGTGATGTAAGTAAGTGTGCAATTAACAGCAACATG 1008
 DB 781 ACTTTGGCAGAAAGCAAGTGTGAGTGTGATGTAAGTAAGTGTGCAATTAACAGCAACATG 840
 QY 1009 GGCAGTGAATTCAGGACCATGGAGATGATTCAGATGAAGAAATGTGCGACAGACACA 1068
 DB 841 GGCAGTGAATTCAGGACCATGGAGATGATTCAGATGAAGAAATGTGCGACAGACACA 900
 QY 1069 TCCCGGAGCCAGAACTCCAGCTAGGCTTACCAATGAAAGTTGCCAGCCAGCTTG 1128
 DB 901 TCCCGGAGCCAGAACTCCAGCTAGGCTTACCAATGAAAGTTGCCAGCCAGCTTG 960
 QY 1129 GAAGGGAAGAAAT 1188
 DB 961 GAAGGGAAGAAAT 1020
 QY 1189 TACATTCGCCAAGATCACTTAGACAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGTATA 1248
 DB 1021 TACATTCGCCAAGATCACTTAGACAAAGAAAGAAAGAAAGAAAGAAAGAAAGTATA 1080
 QY 1249 GTTCTTGTCAATTAAGTACTGCTAGTTGAACAGCTCTCCGAGAGAGTTCACCAATT 1308
 DB 1081 GTTCTTGTCAATTAAGTACTGCTAGTTGAACAGCTCTCCGAGAGAGTTCACCAATT 1140
 QY 1309 TTGAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1368
 DB 1141 TTGAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
 QY 1369 CCAGAGTTGTCAAGTCTGTGATATATATATATATATATATATATATATATATATATAT 1428
 DB 1201 CCAGAGTTGTCAAGTCTGTGATATATATATATATATATATATATATATATATATATAT 1257
 QY 1429 CTCTTAACCTTGGAAATGAGAAAGATGCTGTGTTCAATTTGTCAAGCTTTTCTCTATT 1488
 DB 1258 CTCTTAACCTTGGAAATGAGAAAGATGCTGTGTTCAATTTGTCAAGCTTTTCTCTATT 1317
 QY 1489 ATCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1548
 DB 1318 ATCATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1377
 QY 1549 TATTGATGACAGAGTGTGAAAGCAATAGACTCAAGAAAGAAAGAAAGCAATGATTTC 1608
 DB 1378 TATTGATGACAGAGTGTGAAAGCAATAGACTCAAGAAAGAAAGCAATGATTTC 1437
 QY 1609 CTTCCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1668
 DB 1438 CTTCCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1497
 QY 1669 AAAGCTGAAGAAACATTTTAAAGTATGTCGAATCTTGAATGATTAATTAATTAATTA 1728
 DB 1498 AAAGCTGAAGAAACATTTTAAAGTATGTCGAATCTTGAATGATTAATTAATTAATTA 1557
 QY 1729 GTTAAGAAAGAAACCTTGATCACTGAAAAACCAATACAGAGAGCAATGCAAGAAATTTGGCC 1788
 DB 1558 GTTAAGAAAGAAACCTTGATCACTGAAAAACCAATACAGAGAGCAATGCAAGAAATTTGGCC 1617

QY 1789 ATTGCAGATGACACAGAGAGATCATTTAAAGAAACTTGTAGAAATATGACACAGG 1848
 DB 1618 ATTGCAGATGACACAGAGAGATCATTTAAAGAAACTTGTAGAAATATGACACAGG 1677
 QY 1849 ATTCAAACTTATGTCAAAATGAGTCCATGTGAGATTTTGGAACTCAATATGACACA 1908
 DB 1678 ATTCAAACTTATGTCAAAATGAGTCCATGTGAGATTTTGGAACTCAATATGACACA 1737
 QY 1909 TGGGCATTCATTAATGAAAAAAGCTGCAAAAAAGAAATGCAAAATGCTGTGTTG 1968
 DB 1738 TGGGCATTCATTAATGAAAAAAGCTGCAAAAAAGAAATGCAAAATGCTGTGTTG 1797
 QY 1969 GCAGAACATTTAGAGAAATGACATGAGCCCTACAAATTAATGACACAAATGAAATGATA 2028
 DB 1798 GCAGAACATTTAGAGAAATGACATGAGCCCTACAAATTAATGACACAAATGAAATGATA 1857
 QY 2029 GAT--GCGTATACATCTTGAACCTTCTATATGAGAGAGAAAGATTAATAGTTTGCAG 2086
 DB 1858 GATCCGCGTATACATCTTGAACCTTCTATATGAGAGAGAAAGATTAATAGTTTGCAG 1917
 QY 2087 TCATGAAATGATGATGATGATGAGGCGGTGATGATGATGATGATGATGATGATGATG 2146
 DB 1918 TCATGAAATGATGATGATGATGAGGCGGTGATGATGATGATGATGATGATGATGATG 1977
 QY 2147 AGCATGATTTAAAGAACTTTGAAACTGATGAAACAGATAGATTCCTGACATTAAT 2206
 DB 1978 AGCATGATTTAAAGAACTTTGAAACTGATGAAACAGATAGATTCCTGACATTAAT 2037
 QY 2207 TTTTTCAAAAAATAAATGTTGAAAAAGGCTGCTGAAAAACCAATAATGAAATGAAA 2266
 DB 2038 TTTTTCAAAAAATAAATGTTGAAAAAGGCTGCTGAAAAACCAATAATGAAATGAAA 2097
 QY 2267 AGCTGACCAATTAAGAAATACCATTAATGAGCAATATCTAGAGATGATGATGATGATG 2326
 DB 2098 AGCTGACCAATTAAGAAATACCATTAATGAGCAATATCTAGAGATGATGATGATGATG 2157
 QY 2327 GAGGATTAATCTTTACAAAAACAGACAGATGATGATGATGATGATGATGATGATG 2386
 DB 2158 GAGGATTAATCTTTACAAAAACAGACAGATGATGATGATGATGATGATGATGATG 2217
 QY 2387 AAAATGAAAAATTTGCTGAGTAGTAGAGTAAAGCCCATCTGATGATGATGATGATG 2446
 DB 2218 AAAATGAAAAATTTGCTGAGTAGTAGAGTAAAGCCCATCTGATGATGATGATGATG 2277
 QY 2447 GCAAGTAGTTCAAAACCCAGACAGATGATGATGATGATGATGATGATGATGATG 2506
 DB 2278 GCAAGTAGTTCAAAACCCAGACAGATGATGATGATGATGATGATGATGATGATG 2337
 QY 2507 CTGGAATAATCAATCTGCTTATGCTTACCAAGTGGCAGAGAGAGTCTGATATTAATG 2566
 DB 2338 CTGGAATAATCAATCTGCTTATGCTTACCAAGTGGCAGAGAGAGTCTGATATTAATG 2397
 QY 2567 AATGTAACATTTGATCGCTTATGCTGCTGACCAATGAATAGCCATGCTCCAGGCC 2626
 DB 2398 AATGTAACATTTGATCGCTTATGCTGCTGACCAATGAATAGCCATGCTCCAGGCC 2457
 QY 2627 GTGGTGAGCAGAGAGTATGATGAGAGACCTAGCTGCTGCTGCTGCTGCTGCTGCTG 2686
 DB 2458 GTGGTGAGCAGAGAGTATGATGAGAGACCTAGCTGCTGCTGCTGCTGCTGCTGCTG 2517
 QY 2687 TTATGCAATGAGAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 2746
 DB 2518 TTATGCAATGAGAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 2577
 QY 2747 GTGCTTAAAAATGAAACAGAGAGTATGCTCATATGAAATTTGGAATTTGAGATGCAA 2806
 DB 2578 GTGCTTAAAAATGAAACAGAGAGTATGCTCATATGAAATTTGGAATTTGAGATGCAA 2637
 QY 2807 GTATATGAAAAAGAAATGAAACAGAGAGTATGCTCATATGAAATTTGGAATTTGAGAT 2866
 DB 2638 GTATATGAAAAAGAAATGAAACAGAGAGTATGCTCATATGAAATTTGGAATTTGAGAT 2697
 QY 2867 CATCACTAATTAATCTTCTTGTGCAAAAAAGCTGAGTGTGCTAGCCCTTCTCTGAGAGATA 2926

DB 2698 CATCACTAATTAATCTTCTTGTGCAAAAAAGCTGAGTGTGCTAGCCCTGCTCTGAGAGATA 2757
 QY 2927 TCCATGTAATGAGAAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2986
 DB 2758 TCCATGTAATGAGAAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2817
 QY 2987 TTGTAGAGAAACCAAAACAGCTGCAAAAGAGTGGCCGATCATCAATTAATGATGAAA 3046
 DB 2818 TTGTAGAGAAACCAAAACAGCTGCAAAAGAGTGGCCGATCATCAATTAATGATGAAA 2877
 QY 3047 TCATCTGCAATGTTGGCCAGGCTTGGGACAAATGATGATGATGATGATGATGATGATG 3106
 DB 2878 TCATCTGCAATGTTGGCCAGGCTTGGGACAAATGATGATGATGATGATGATGATGATG 2937
 QY 3107 CTGTCTCAAAATTAAGGAATTTTGTAGTGTGTTTCAAAAAATTAATTAACAAACAAAT 3166
 DB 2938 CTGTCTCAAAATTAAGGAATTTTGTAGTGTGTTTCAAAAAATTAATTAACAAACAAAT 2997
 QY 3167 ACAAAGAGGAGTACATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 3226
 DB 2998 ACAAAGAGGAGTACATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 3057
 QY 3227 TATTTAGTATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3286
 DB 3058 TATTTAGTATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3117
 QY 3287 ATTAAATATGATTA 3300
 DB 3118 ATTAAATATGATTA 3131

RESULT 4
 AAS40960
 ID AAS40960 standard; cDNA: 1967 BP.
 XX
 AC AAS40960:
 XX
 DT 17-DEC-2001 (first entry)
 XX
 DE cDNA encoding novel human enzyme polypeptide #176.
 KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
 KW lysase; hyperproliferative disorder; immunodeficiency disorder;
 KW autoimmune disorder; neurological disorder; metabolic disorder;
 KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
 KW blood-related disorder; infectious disorder; gene therapy; cytostatic;
 KW anti arthritic; nephrotropic; anticoagulant; ss.
 OS Homo sapiens.
 PM WO20015301-A2.
 XX
 PD 02-AUG-2001.
 PF 17-JAN-2001; 2001WO-US01239.
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198122.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 14-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226686.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0233399.
 PR 14-SEP-2000; 2000US-0233400.
 PR 14-SEP-2000; 2000US-0233401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234224.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0244674.

PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246529.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246537.
 PR 08-NOV-2000; 2000US-0246539.
 PR 08-NOV-2000; 2000US-0246569.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249266.
 PR 17-NOV-2000; 2000US-0249267.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0251989.
 PR 06-DEC-2000; 2000US-0251479.
 PR 06-DEC-2000; 2000US-0251856.
 PR 06-DEC-2000; 2000US-0251857.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI: 2001-465566/50.
P-PSDB: AAU23090.

Novel polypeptides and polynucleotides useful for diagnosing, preventing, treating neural, immune system, muscular, reproductive, pulmonary, cardiovascular, renal, proliferative disorders and cancerous diseases

Claim 4; SEQ ID No 186; 1180pp; English.

The present invention relates to the isolation of novel human enzyme polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences encoding them. The enzyme polypeptides of the invention may comprise the functional classes of oxidoreductases, transferases, hydrolases, lyases, isomerases or ligases. The sequences of the invention are useful in the diagnosis, treatment, prevention and/or prognosis of a wide range of disorders including hyperproliferative disorders (e.g. cancer), immunodeficiency disorders (e.g. AIDS) autoimmune disorders (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease), metabolic disorders (e.g. phenylketonuria), inflammatory disorders (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),

CC blood-related disorders (e.g. haemophilia), reproductive disorders
 CC (e.g. infertility) and infectious disorders (e.g. Influenza). The
 CC polynucleotides of the invention can also be used in gene therapy.
 CC AA40785-AA41684 represent cDNA sequences encoding for the novel human
 CC enzyme polypeptides of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences

XX Sequence 1967 BP: 723 A; 344 C; 416 G; 476 T; 8 other:

Query Match 55.8%; Score 1878; DB 22; Length 1967;
 Best Local Similarity 98.5%; Pred. No. 0;
 Matches 1941; Conservative 5; Mismatches 18; Indels 7; Gaps 5;

Oy 1047 AGAGATGTGGCAGCAGAGCATCCCGAGCCAGACCTCCAGCTCAGGCTTACCAAT 1106
 Db 2 AGAGATGTGGCAGCAGAGCATCCCGAGCCAGACCTCCAGCTCAGGCTTACCAAT 61
 Oy 1107 GGAATGCGCCAGCGCTGGAGAGAGATATCATCATCTGCTCTCTACAGGGAG 1166
 Db 62 GGAATGCGCCAGCGCTGGAGAGAGATATCATCATCTGCTCTCTACAGGGAG 121
 Oy 1167 TGGAAAAACAGAGTGGCTGTTCATTGCCAAGATCACTTAGACAGAGAAAAAGC 1226
 Db 122 TGGAAAAACAGAGTGGCTGTTCATTGCCAAGATCACTTAGACAGAGAAAAAGC 181
 Oy 1227 ATCTGAGCTGGAAAAATTATAGTTCTGTCTCAATAGTACTGCTAGTGAACGCTCTT 1286
 Db 182 ATCTGAGCTGGAAAAATTATAGTTCTGTCTCAATAGTACTGCTAGTGAACGCTCTT 241
 Oy 1287 CCGAAGAGTTCACACCATTTTGAAGAAATGATGCTGTTATGATTAAGTGGTGA 1346
 Db 242 CCGAAGAGTTCACACCATTTTGAAGAAATGATGCTGTTATGATTAAGTGGTGA 301
 Oy 1347 TACCACTGAAAAATATATTCAGAAAGTTCAGAGTCTGTATTTATATCACTAG 1406
 Db 302 TACCACTGAAAAATATATTCAGAAAGTTCAGAGTCTGTATTTATATCACTAG 361
 Oy 1407 AGCTAAATCTTGAAGAACTCCCTCTTAACCTTGAAAAAGAGAGAGTCTGTCTCA 1466
 Db 362 AGCTAAATCTTGAAGAACTCCCTCTTAACCTTGAAAAAGAGAGAGTCTGTCTCA 421
 Oy 1467 ATTGTGAGCTTTCCCTCATATCATGATGATGATGATGATGATGATGATGATGAT 1526
 Db 422 ATTGTGAGCTTTCCCTCATATCATGATGATGATGATGATGATGATGATGATGAT 481
 Oy 1527 GTATATATCATGAGGATTTATTTGATGAGAGAGTGAAGAAACAATAGACTAGAA 1586
 Db 482 GTATATATCATGAGGATTTATTTGATGAGAGAGTGAAGAAACAATAGACTAGAA 541
 Oy 1587 AGAAAAACAACAGATGATTCCTCTCTCAGATCTGAGACTAGAGCTTCACTGGTGT 1646
 Db 542 AGAAAAACAACAGATGATTCCTCTCTCAGATCTGAGACTAGAGCTTCACTGGTGT 601
 Oy 1647 TGGAGGGGCGCAGAGCAAGCCAAAGCTGAACACATTTTAAAGCTATGCGCAACT 1706
 Db 602 TGGAGGGGCGCAGAGCAAGCCAAAGCTGAACACATTTTAAAGCTATGCGCAACT 661
 Oy 1707 TGATGATTTACTATTAAACTGTTAAGAAAAACCTGTATCACTGAGAAAAACAATACA 1766
 Db 662 TGATGATTTACTATTAAACTGTTAAGAAAAACCTGTATCACTGAGAAAAACAATACA 721
 Oy 1767 GGAGGCATCAAGAAAGTTGGCATTGACAGATCAACACAGAGAGATCAATTAAGAGAA 1826
 Db 722 GGAGGCATCAAGAAAGTTGGCATTGACAGATCAACACAGAGAGATCAATTAAGAGAA 781
 Oy 1827 ACTTCTAGAAATATGACAGAGATCAAACTTATGTCAATAGATCAAACTTCAGATTT 1886
 Db 782 ACTTCTAGAAATATGACAGAGATCAAACTTATGTCAATAGATCAAACTTCAGATTT 841
 Oy 1887 TGGAACTCAACCTATGAACAATGGGCACTTCAATGAAAAAAGCTCTAAAAAAGG 1946
 Db 1946 TGGAACTCAACCTATGAACAATGGGCACTTCAATGAAAAAAGCTCTAAAAAAGG 1946

Db 842 TGAAGTCAACCTATGACAAATGGGCACTTCAATGAAAAAAGCTGCAAAAGAGG 901
 Oy 1947 AAATCGAAGAGAGCTGTTGTGCGAAGCATTTTGAAGATGACATAGAGCTTACAAT 2006
 Db 902 AAATCGAAGAGAGCTGTTGTGCGAAGCATTTTGAAGATGACATAGAGCTTACAAT 961
 Oy 2007 TAATGACCAATTTGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2066
 Db 962 TAATGACCAATTTGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1021
 Oy 2067 GAAAGATAGAAAGTTTGCAGTCAAGAGATGATGATGATGATGATGATGATGATGATGAT 2126
 Db 1022 GAAAGATAGAAAGTTTGCAGTCAAGAGATGATGATGATGATGATGATGATGATGATGAT 1081
 Oy 2127 TTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2186
 Db 1082 TTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1141
 Oy 2187 TAGATTTCTCATGCTTATTTTGAAGAAATGATGATGATGATGATGATGATGATGATGAT 2246
 Db 1142 TAGATTTCTCATGCTTATTTTGAAGAAATGATGATGATGATGATGATGATGATGATGAT 1201
 Oy 2247 CCGAAGATATGAAAAATGAAAAAGCTGACCAAAATTAAGAAATACATTAATGAAATATAATAC 2306
 Db 1202 CCGAAGATATGAAAAATGAAAAAGCTGACCAAAATTAAGAAATACATTAATGAAATATAATAC 1261
 Oy 2307 TAGAGCTGAGAAATCAGCAGAGAGATTAATTTTCAAAAAACAGACAGAGTGCATATGC 2366
 Db 1262 TAGAGCTGAGAAATCAGCAGAGAGATTAATTTTCAAAAAACAGACAGAGTGCATATGC 1321
 Oy 2367 GCTTCCCGAGTGAATTAAGTGAATGAAAAATTTGCTGAATAGAGTCAAAATTCAGCA 2426
 Db 1322 GCTTCCCGAGTGAATTAAGTGAATGAAAAATTTGCTGAATAGAGTCAAAATTCAGCA 1381
 Oy 2427 TGTGATGAGAGCTGACACAGCAGCTGATGATGATGATGATGATGATGATGATGATGAT 2486
 Db 1382 TGTGATGAGAGCTGACACAGCAGCTGATGATGATGATGATGATGATGATGATGATGAT 1441
 Oy 2487 AGTCAATTGTAATTTTGGCAGTGGAAAAATCAATCTGTTATGCTTACCAATTCAGCA 2546
 Db 1442 AGTCAATTGTAATTTTGGCAGTGGAAAAATCAATCTGTTATGCTTACCAATTCAGCA 1501
 Oy 2547 AGAAGTGTGATATTAAGAAATGAATCAATGTTATGCTTATGCTTATGCTTATGCT 2606
 Db 1501 AGAAGTGTGATATTAAGAAATGAATCAATGTTATGCTTATGCTTATGCTTATGCT 1560
 Oy 2607 AATAGCCATGGTCCAGCGCCGCTGCTGAGCCAGAGCTGATGAGACACATATCTGCT 2666
 Db 1561 AATAGCCATGGTCCAGCGCCGCTGAGCCAGAGCTGATGAGACACATATCTGCT 1617
 Oy 2667 TGTCTACAGTGGTTCAGAGTTATTCGAACATGAGACAGTTAATGATTTCCGATTAAGAT 2726
 Db 1617 TGTCTACAGTGGTTCAGAGTTATTCGAACATGAGACAGTTAATGATTTCCGATTAAGAT 1677
 Oy 2727 GATGATTAAGAGTATACATTTGTTCAAAATATGAACAGAGAGATATGCTTAAGAT 2786
 Db 1678 GATGATTAAGAGTATACATTTGTTCAAAATATGAACAGAGAGATATGCTTAAGAT 1744
 Oy 2787 TTTGGAATTAAGATGAAGATTAATGAAAAAGAAAAATGAAAAACAAGAGAAATATTGC 2846
 Db 1735 TTTGGAATTAAGATGAAGATTAATGAAAAAGAAAAATGAAAAACAAGAGAAATATTGC 1794
 Oy 2847 CAAGATTTACAAGATTAACCATCTACTATACTTTCTTTGCAAAAACTGATGATGATGAT 2906
 Db 1795 CAAGATTTACAAGATTAACCATCTACTATACTTTCTTTGCAAAAACTGATGATGATGAT 1854
 Oy 2907 AGCTGTTCTGGGGAAGATATTCATGATGATGATGATGATGATGATGATGATGATGATGAT 2966
 Db 1855 AGCTGTTCTGGGGAAGATATTCATGATGATGATGATGATGATGATGATGATGATGATGAT 1914
 Oy 2967 AGAATTTCAAGAACTTACATTTGTAAGAGAAACAAAGCACTGCAAAAAAGAA 3027
 Db 1915 AGAATTTCAAGAACTTACATTTGTAAGAGAAACAAAGCACTGCAAAAAAGAA 1975

Db 1141 GACCCGAGATTCAGGAACTTACATGTAGAGAAACAAAGACTG-7AAAGAGTG 1200
 QY 3021 TCGCGACTCAATTAATAGTGAATCATCTGCAAAATGCGCCAGGCTTGGGAAACAT 3080
 Db 1201 TCGCGACTCAATTAATAGTGAATCATCTGCAAAATGCGCCAGGCTTGGGAAACAT 1260
 QY 3081 GATGGGCAAAAGGCTTGAATTTGCTGCTCAAA- AAGCAATTTT- AGTGGTTC 3140
 Db 1261 GATGGGCAAAAGGCTTGAATTTGCTGCTCAAA- AAGCAATTTT- AGTGGTTC 1320
 QY 3141 CAAAAATTAATCAACAAAGAAACATATCAAAAAAGTGG-AGAATTACCT- CACATTTCC 3200
 Db 1321 CAAAAATTAATCAACAAAGAAACATATCAAAAAAGTGGAGATTACCT- CACATTTCC 1380
 QY 3201 CAATCTGACTATGCAATGCTGTTTATTAGTGTAGTGAAGATTAGCACT- GATTGAA 3260
 Db 1381 CAATCTGACTATGCAATGCTGTTTATTAGTGTAGTGAAGATTAGCACT- GATTGAA 1440
 QY 3261 TTTCTTTAAATATCTATGCTTAAACATTTAATGATGATGATGATGATGATGATG 3320
 Db 1441 TTTCTTTAAATATCTATGCTTAAACATTTAATGATGATGATGATGATGATGATG 1500
 QY 3321 CTACAGACTGACATTAAGAAATCAATAAATGATTTT- ACTCTG 3385
 Db 1501 CTACAGACTGACATTAAGAAATCAATAAATGATTTT- ACTCTG 1511

RESULT 6

AB04916
ID ABA04916 standard; DNA; 1443 BP.

XX ABA04916;

DT 05-MAR-2002 (first entry)

XX Human RNA helicase RH16 related DNA sequence #3.

XX Human: RH16; RNA helicase; cytosolic; vireonide; anti-HIV;

XX Immunosuppressive; immunostimulatory; antileukemic; a tharthritis;

XX antileukosclerotic; osteopathic; antidiabetic; hepatotropic;

XX antileukosclerotic; cancer; infection; HIV; hepatitis; genetic disease;

XX autoimmune disease; graft rejection; vaccine; ds.

XX Homo sapiens.

XX WO200185955-A1.

XX 15-NOV-2001.

XX 11-MAY-2001: 2001MO-FR01441.

XX 11-MAY-2000: 2000FR-0006030.

XX (ISTA-) ISTAC.

XX (INST) INST PASTEUR LILLE.

XX Bahr G, Cocude C, Capron A;

XX WPI: 2002-082898/11.

XX New polypeptide, useful for treating and diagnosing cancer or

XX inflammation, and drug screening, comprises a human polynucleotide

XX homologous to RNA helicase

XX Example 2; Page 94-95; 114pp; French.

XX The present invention relates to human RH16 (see ABA04916). RH16 is a

XX 116kDa protein and has homology to RNA helicases (DEXH box). RH16 and

XX its coding sequence are useful for treating cancer, acute or chronic

XX infections (especially by HIV or hepatitis B or C); inherited genetic

XX diseases; (auto)immune diseases (particularly rheumatoid arthritis,

XX arteriosclerosis, osteoporosis and diabetes, but many others listed) and

CC to prevent graft rejection. RH16 and its coding sequence are also useful
 CC for inducing, or increasing, the immune response to a vaccine. The
 CC present sequence was used in an example from the present invention.

XX Sequence 1443 BP; 508 A; 261 C; 313 G; 360 T; 1 other:

Query Match 42.7%; Score 1435.6; DB 24; Length 1443;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 1438; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 687 CAAAGAAAACCTGCTCTGCTGCAATTTCTCAATGTTCTTGTCAACAGCAAA- AAG-CT 746
 Db 1 CAAAGAAAACCTGCTCTGCTGCAATTTCTCAATGTTCTTGTCAACAGCAAA- AAG-CT 60
 QY 747 TGTCCAGAGTAAAGAGCTGCTGATGCTGCAAAAGCAATGAGAGATGAA-ATTTC 806
 Db 61 TGTCCAGAGTAAAGAGCTGCTGATGCTGCAAAAGCAATGAGAGATGAA-ATTTC 120
 QY 807 ACAAGTTGATGCTGCTCAAGTGAAGAGCAACTTCTTCAACCAAGTTCAG-CAAAATCT 866
 Db 121 ACAAGTTGATGCTGCTCAAGTGAAGAGCAACTTCTTCAACCAAGTTCAG-CAAAATCT 180
 QY 867 GGAGAAAGAGCTGCTGGGCTAGGAGAAATACATCATCAATATCTTTTGCAG-ATCTC 926
 Db 181 GGAGAAAGAGCTGCTGGGCTAGGAGAAATACATCATCAATATCTTTTGCAG-ATCTC 240
 QY 927 TGTAGTTCAAGATGAGACACAAAGTTTGGCAAGAGAGTGCAGCTGCTTGA-AGAAAG 986
 Db 241 TGTAGTTCAAGATGAGACACAAAGTTTGGCAAGAGAGTGCAGCTGCTTGA-AGAAAG 300
 QY 987 TCTTGACATAACAGCAACATGCGGAGTATTCAGGACCAATGGGAAGTATTCAG-ATCT 1046
 Db 301 TCTTGACATAACAGCAACATGCGGAGTATTCAGGACCAATGGGAAGTATTCAG-ATCT 360
 QY 1047 AGAGAAATGCGGACCAAGAGCATCTCCGAGGCCAATCTCCAGCTCAGGCC-CTA-CAAAAT 1106
 Db 361 AGAGAAATGCGGACCAAGAGCATCTCCGAGGCCAATCTCCAGCTCAGGCC-CTA-CAAAAT 420
 QY 1107 GGAAGTGGCCGAGCCAGCTTGGAAAGGGAAGAAATTCATCATCTGCTCCCT-ATCGAGGAG 1166
 Db 421 GGAAGTGGCCGAGCCAGCTTGGAAAGGGAAGAAATTCATCATCTGCTCCCT-ATCGAGGAG 480
 QY 1167 TGGAAAAACCAAGAGTGGCTGTTTACATTTGCCAAGATCACTTAGCAAGAA-CAAAAAGAG 1226
 Db 481 TGGAAAAACCAAGAGTGGCTGTTTACATTTGCCAAGATCACTTAGCAAGAA-CAAAAAGAG 540
 QY 1227 ATCTGAGCTGGAAGAAATATAGTTCTGTCAATTAAGTACTGCTAGTTGA-ATCTCTT 1286
 Db 541 ATCTGAGCTGGAAGAAATATAGTTCTGTCAATTAAGTACTGCTAGTTGA-ATCTCTT 600
 QY 1287 CCGCAAGAGCTTCCAAACATTTTGAAGAAATGATGCTTATTTGATTAAG-AGTGA 1346
 Db 601 CCGCAAGAGCTTCCAAACATTTTGAAGAAATGATGCTTATTTGATTAAG-AGTGA 660
 QY 1347 TACCAACTGAAATATATCTTCCAGAAAGTGTCAAGTCTGTATATAT-ATACAG 1406
 Db 661 TACCAACTGAAATATATCTTCCAGAAAGTGTCAAGTCTGTATATAT-ATACAG 720
 QY 1407 AGCTCAATCTTGAAGAACTCCCTCTTAAACTTGAAGAAATGAGAAAGTGTCT-ATCTCA 1466
 Db 721 AGCTCAATCTTGAAGAACTCCCTCTTAAACTTGAAGAAATGAGAAAGTGTCT-ATCTCA 780
 QY 1467 ATTGTGAGCTTTTCCCTCATATATATGATTAAGTATGATCAACCAAG-CAAGAGCT 1526
 Db 781 ATTGTGAGCTTTTCCCTCATATATATGATTAAGTATGATCAACCAAG-CAAGAGCT 840
 QY 1527 GTATATATACATGAGAGCATTTATTTGATGAGAAAGTGAAGAAATAGAA-AT-AGCA 1586
 Db 841 GTATATATACATGAGAGCATTTATTTGATGAGAAAGTGAAGAAATAGAA-AT-AGCA 900
 QY 1587 AGAAAAACAAGAGTATTCCTCTCTCAGTACTGGAGCTCAAGAGCTTCACT-CTCTCT 1646
 Db 901 AGAAAAACAAGAGTATTCCTCTCTCAGTACTGGAGCTCAAGAGCTTCACT-CTCTCT 960

XX	Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Kou F, Wang D;
Pt	Meng J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
Pt	Zhao QA, Zhou P, Goodrich R, Dimanac RT;
XX	
DR	WPI: 2001-442253/47.
DR	P-P5DB; AAM41915.
XX	
PT	Novel nucleic acids and polypeptides, useful for treating disorders
PT	such as central nervous system injuries -
XX	
PS	Claim 1: SEQ ID NO 5060; 10078bp; English.
CC	
CC	The invention relates to human nucleic acids (AA157798-AA161369) and
CC	the encoded polypeptides (AAM38642-AAM42213) with neurotropic,
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful
CC	in gene therapy. A composition containing a polypeptide or polynucleotide
CC	of the invention may be used to treat diseases of the peripheral nervous
CC	system, such as peripheral nervous injuries, peripheral neuropathy and
CC	localised neuropathies and central nervous system diseases, such as
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC	utilisation of the activities such as: Immune system suppression,
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC	assays for receptor activity, arthritis and inflammation, leukaemia and
CC	C.N.S disorders.
CC	Note: The sequence data for this patent did not form part of the patented
CC	specification.
XX	
SQ	Sequence 1382 BP; 513 A; 217 C; 285 G; 367 T; 0 other:
	Query Match 39.1%; Score 1316.6; DB 22; Length 1382;
	Best Local Similarity 99.1%; Pred. No. 0;
	Matches 1355; Conservative 0; Mismatches 9; Indels 3; Gaps 3;
OY	2002 CAATTAATGACACACTTGGATGATAGATGGGTACTACTCATCTTGAACTCGATAAT 2061
Dd	13 CAATTAAAGCACCAATTCGAATGATAGATGGGTACTACTCATCTTGAACTTCTATAAT 72
OY	2062 GAAGAGAAAAGATAGAAGTTTGCAGTCATAGAAGATGATAGTAGAGAGGTCTATGAT 2121
Dd	73 GAAGAGAAAAGATAGAAGTTTGCAGTCATAGAAGATGATAGTAGAGAGGTCTATGAT 132
OY	2122 GAGTATTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2181
Dd	133 GAGTATTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 192
OY	2182 ACAGATA-CATTGTCATCATTCTTTTTTAAAAACAATTAAT-GTGGAAAACCTGG 2239
Dd	193 ACAGATACGATTTCATCATTCTTTATTTTTTAAAAACAATTAATAGTGAATAAGCTGG 252
OY	2240 CTGAAACCAGCAATATGAAAATGAAAAGCTGACCAATTAAGAAAATACCAATAGAG 2299
Dd	253 CTGAAACCAGCCCTAATATGAAAATGAAAAGCTGACCAATTAAGAAAATACCAATAGAG 312
OY	2300 AATATACTAGAGACTGAGAAATCAGACAGAGGAATATCTTTACAAAAACACAGATAGAG 2359
Dd	313 AATATACTAGAGACTGAGAAATCAGACAGAGGAATATCTTTACAAAAACACAGAGAGAG 372
OY	2360 CATATGCCGTTCCAGTGGATTTACTGAAAATGAAAAAATTTCTGTAAGTAGAGCAAG 2419
Dd	373 CATATGCCGTTCCAGTGGATTTACTGAAAATGAAAAAATTTCTGTAAGTAGAGCAAG 432
OY	2420 CCCACATCTGATTTGAGCTGGACACAGCAGTGAAGTCAAACCCATGACACATAATGAAT 2479
Dd	433 CCCACATCTGATTTGAGAGCTGGACACAGCAGTGAAGTCAAACCCATGACACATAATGAAT 492
OY	2480 AAAAAGAAAGTCAATGTAATTAATTTGCACAGGAAAAATCAATCGTATACG-TATACAG 2539
Dd	493 AAAAAGAAAGTCAATGTAATTAATTTGCACAGGAAAAATCAATCGTATACG-TATACAG 552
OY	2540 TGCGAAGAAAGGTGTGATTAATTAAGAAATGAAACATGTTATCGGTATGATCTCTGTA 2599

QY 2643 TGATGAGACACCTACGTCCTGGTGTGTCACAGTGTGTCAGAGATTATGTAACATGAGAC 2702
DB 541 TGATGAGACACCTACGTCCTGGTGTGTCACAGTGTGTCAGAGATTATGTAACATGAGAC 600
QY 2703 AGTTATGATTTCCGAGAGAAATGATGATTAAGCTATACATGCTTGAATAATATGAA 2762
DB 601 AGTTATGATTTCCGAGAGAAATGATGATTAAGCTATACATGCTTGAATAATATGAA 660
QY 2763 ACCAGAGGATGCTCATAGATTTTGGAAATTTACAGATGCAAGATATGAGAAAGAA 2822
DB 661 ACCAGAGGATGCTCATAGATTTTGGAAATTTACAGATGCAAGATATGAGAAAGAA 720
QY 2823 AATGAAACCAAGAGAAATATTTGCCAAGCATTTACAGAAATACCCATCAATAATAC 2882
DB 721 AATGAAACCAAGAGAAATATTTGCCAAGCATTTACAGAAATACCCATCAATAATAC 780
QY 2883 CATTTCACAAACCTGAGGTGTGACCTGTGTCGAGGAGAAATGCAATCATTAATGATTA 2942
DB 781 CATTTCACAAACCTGAGGTGTGACCTGTGTCGAGGAGAAATGCAATCATTAATGATTA 840
QY 2943 AATGATCAGCTCATATGATGACCCAGAAATTTCAAGAACTTACATTTGTAAGAAACAA 3002
DB 841 AATGATCAGCTCATATGATGACCCAGAAATTTCAAGAACTTACATTTGTAAGAAACAA 900
QY 3003 AGCAGCTGCAAAAGAGTGTGCGGACTATGCAATTAATGTTGTAATGATGCTG 3062
DB 901 AGCAGCTGCAAAAGAGTGTGCGGACTATGCAATTAATGTTGTAATGATGCTG 960
QY 3063 CCAGGCTGGGAGCAATGATGTCGCAAAAGGCTGATGCTGTCGCAAAATAG 3122
DB 961 CCAGGCTGGGAGCAATGATGTCGCAAAAGGCTGATGCTGTCGCAAAATAG 1020
QY 3123 GAATTTGTAGTGTGTTTCAAAATATATTCACAAAGAAACAAATAGAGGAGTACA 3182
DB 1021 GAATTTGTAGTGTGTTTCAAAATATATTCACAAAGAAACAAATAGAGGAGTACA 1080
QY 3183 ATTACCTATCAGATTTCCCAATCTTGACATTCAGATTCCTGTTTAAATGATGAGAGA 3242
DB 1081 ATTACCTATCAGATTTCCCAATCTTGACATTCAGATTCCTGTTTAAATGATGAGAGA 1140
QY 3243 TTAGACCTGATGAGATTTCTTTAAATACATGATTAACATTAATGATGATG 3302
DB 1141 TTAGACCTGATGAGATTTCTTTAAATACATGATTAACATTAATGATGATG 1200
QY 3303 ATTATGATTTGATTTGCTGACAGAACTGACATTAAGATCAATTAATGATGATG 3362
DB 1201 ATTATGATTTGATTTGCTGACAGAACTGACATTAAGATCAATTAATGATGATG 1260
QY 3363 CT 3364
DB 1261 NT 1262

RESULT 9
AAS41517
ID AAS41517 standard: cDNA: 1319 BP.
XX
AC AAS41517:
XX
DT 17-DEC-2001 (first entry)
XX
DE cDNA encoding novel human enzyme polypeptide #733.
XX
KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; gene therapy; cytostatic;
KW anti-arthritis; nephrotropic; anticoagulant; ss.
XX
OS Homo sapiens.
XX

PN W020015301-A2.
XX 02-AUG-2001.
PD 17-JAN-2001; 2001W0-US01239.
XX
PE
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232337.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0233065.
PR 14-SEP-2000; 2000US-0233066.
PR 14-SEP-2000; 2000US-0233067.
PR 14-SEP-2000; 2000US-0233068.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.

PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0244647.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 06-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259676.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SW;

XX
 DR WPI: 2001-465566/50.
 DR P-PSDB: AAU23647.
 XX
 PT Novel polypeptides and polynucleotides useful for diagnosis;
 PT preventing, treating neural, immune system, muscular, reproductive,
 PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous
 PT diseases
 XX
 PS Claim 4: SEQ ID NO 743; 1180pp; English.
 XX
 CC The present invention relates to the isolation of novel human enzyme
 CC polypeptides (AAU22915-AAU23184), and the cDNA and genomic sequences
 CC encoding them. The enzyme polypeptides of the invention comprise the
 CC functional classes of oxidoreductases, transferases, hydrolases, lyases,
 CC isomerases or ligases. The sequences of the invention are useful in the
 CC diagnosis, treatment, prevention and/or prognosis of a wide range of
 CC disorders including hyperproliferative disorders (e.g. cancer),
 CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders
 CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
 CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders
 CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
 CC blood-related disorders (e.g. haemophilia), reproductive disorders
 CC (e.g. infertility) and infectious disorders (e.g. influenza). The
 CC polynucleotides of the invention can also be used in gene therapy.
 CC AAU40785-AAU41684 represent cDNA sequences encoding for the novel human
 CC enzyme polypeptides of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SO Sequence 1319 BP; 491 A; 217 C; 294 G; 314 T; 3 other;
 Query Match 33.6%; Score 1130.8; DB 22; Length 415;
 Best Local Similarity 95.3%; Pred. No. 47e-270;
 Matches 1253; Conservative 2; Mismatches 40; Indels 40; Gaps 8;
 QY 1493 TTGATGATGTCATCAGCAACCAAGAGAGTATATATACATGAGATTTATT 1552
 DB 24 TTGATGATGTCATCAGCAACCAAGAGAGTATATATACATGAGATTTATT 83
 QY 1553 TGATGAGAGTGTGAAAAACATATAGCTCAAGAAAAACAAACCACTGAT 1612
 DB 84 TGATGAGAGTGTGAAAAACATATAGCTCAAGAAAAACAAACCACTGAT 143
 QY 1613 CTCAGATCTGGGACTTAACAGCTTACTGCTGTGAGGGGCCAGCAAGCAAG 1672
 DB 144 CTCAGATCTGGGACTTAACAGCTTACTGCTGTGAGGGGCCAGCAAGCAAG 203
 QY 1673 CTGAAGAACACATTTAAACATGTGCAATCTGATGATTTACTTTAAATGTTA 1732
 DB 204 CTGAAGAACACATTTAAACATGTGCAATCTGATGATTTACTTTAAATGTTA 263
 QY 1733 AAGAAACCTTGATCACTGAAAAACCAATATACAGAGCCATGCAAGAGTTC 1792
 DB 264 AAGAAACCTTGATCACTGAAAAACCAATATACAGAGCCATGCAAGAGTTC 323
 QY 1793 CAGATGCAACGAGAGAAATCATTTAAAGAAACTCTGAAATTAATGCAAGATTC 1852
 DB 324 CAGATGCAACGAGAGAAATCATTTAAAGAAACTCTGAAATTAATGCAAGATTC 383
 QY 1853 AAACCTTATTTGCAATGATGCAATGCAATTTGGAAGCAACCCATGATATATAGG 1912
 DB 384 AAACCTTATTTGCAATGATGCAATGCAATTTGGAAGCAACCCATGATATATAGG 443
 QY 1913 CCATTCATATGCAAAAAAGCTGCAAAAAAGAAATGCAAAAGACGTGTATATGACG 1972
 DB 444 CCATTCATATGCAAAAAAGCTGCAAAAAAGAAATGCAAAAGACGTGTATATGACG 503
 QY 1973 AACATTTGAGAGAGTACAAATGAGGCTTACAAATTTAATGACCAATTTGAAATATATG 2032
 DB 504 AACATTTGAGAGAGTACAAATGAGGCTTACAAATTTAATGACCAATTTGAAATATATG 563

[illegible]


```

QY 225 GGCACAGGTGAAATGTACATCCAGGTGGAGCCTGTGCTGGACTACCTGCTTTCTGAC 284
    |||||||
DB 235 GGCACAGGTGAAATGTACATC----- 256
QY 285 TCCAGAGGTGAAGAGACAGATTACAGACAGCTCCCACTCCGGGAATCTCCAGAGT 344
    |||||||
DB 257 --CAGAGGTGAAGAGACAGATTACAGACAGCTCCCACTCCGGGAATCTCCAGAGT 314
QY 345 TGAAGTGGCTGAGACACTTGGAGAGGAGTGGGACCTGGTCTCTCTCTCTCTCTCT 404
    |||||||
DB 315 TGAAGTGGCTGAGACACTTGGAGAGGAGTGGGACCTGGTCTCTCTCTCTCTCTCT 374
QY 405 CGTGGAGGCTCCCGAGAACCGGAGCCTCTGGCCGACCCCTACATCTCTCTCTCTCT 464
    |||||||
DB 375 CGTGGAGGCTCCCGAGAACCGGAGCCTCTGGCCGACCCCTACATCTCTCTCTCTCT 434
QY 465 CACGAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 524
    |||||||
DB 435 CACGAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 494
QY 525 CCTCTCTGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 584
    |||||||
DB 495 CCTCTCTGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 554
QY 585 CGAGGAGAACTGTGACAAATGGAAGACAGAACCGGATCTCTCTCTCTCTCTCTCTCT 644
    |||||||
DB 555 CGAGGAGAACTGTGACAAATGGAAGACAGAACCGGATCTCTCTCTCTCTCTCTCTCT 614
QY 645 AAATGATCAGGTGTAAAGAGAGCTACTAAAAAGATTGTGCAAGAAAGAAATCTGTCT 704
    |||||||
DB 615 AAATGATCAGGTGTAAAGAGAGCTACTAAAAAGATTGTGCAAGAAAGAAATCTGTCT 674
QY 705 TCGATTCTGAGATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 764
    |||||||
DB 675 TCGATTCTGAGATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 734
QY 765 CTCTGATCTCTGAGAAAGCAATGAGAGATGAGATTATACAAAGTCTCTCTCTCTCT 824
    |||||||
DB 735 CTCTGATCTCTGAGAAAGCAATGAGAGATGAGATTATACAAAGTCTCTCTCTCTCT 794
QY 825 AGTGAAGAGCAACTCTTTTCAACCAAGCTTCAAGCCAAATCTGGAGAGAGTAACTCT 884
    |||||||
DB 795 AGTGAAGAGCAACTCTTTTCAACCAAGCTTCAAGCCAAATCTGGAGAGAGTAACTCT 854
QY 885 CATGGAGATTAATCATCATGAGATTCATCTTTGAGAGATTTCTCTCTCTCTCTCTCT 937
    |||||||
DB 855 CATGGAGATTAATCATCATGAGATTCATCTTTGAGAGATTTCTCTCTCTCTCTCTCT 907

RESULT 11
AAS01151
ID AAS01151 standard; DNA: 1270 BP.
XX
AC AAS01151;
XX
XX 12-SEP-2001 (first entry)
XX
DE Interferon induced nucleic acid, IFN6.
XX
XX Interferon induced nucleic acid; autoimmune disease; lupus erythematosus;
XX immunodeficiency; stroke; acquired immunodeficiency syndrome; AIDS;
XX graft rejection; viral infection; hepatitis aplastic anaemia; cancer;
XX human immunodeficiency virus; HIV; immune-mediated glomerulonephritis;
XX haematologic disease; chronic neutropenia; myocardial infarction;
XX neurological disease; Alzheimer's disease; Parkinson's disease; tumour;
XX amyotrophic lateral sclerosis; spinal muscular atrophy; human; IFN6; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 1..927
XX FT /tag="a
XX FT /product="IFN6"
XX FT

```

```

PT /note="Interferon induced polypeptide"
XX
XX W0200118208-A2.
XX
XX 15-MAR-2001.
XX
XX 08-SEP-2000; 2000WO-US24704.
XX
XX 08-SEP-1999; 99US-0152921.
XX 20-OCT-1999; 99US-0160575.
XX 20-JAN-2000; 2000US-0177104.
XX 07-SEP-2000; 2000US-0656633.
XX
XX (CURA-) CURAGEN CORP.
XX (BIOJ ) BIOGEN INC.
XX
XX Peyman JA, Da Silva A, Hochman P, Hsu A:
XX
XX WPI: 2001-235201/24.
XX P-PSDB: AMU00298.
XX
XX New interferon induced polypeptides and polynucleotides, useful for the
XX diagnosis, prevention and treatment of immunological, cell
XX proliferative disorders, such as lupus erythematosus, cancer, stroke
XX and Alzheimer's disease.
XX
XX Claim 9; Page 40-43; 134pp; English.
XX
XX The sequence represents interferon induced nucleic acid, IFN6. IFN
XX nucleic acids and polypeptides are useful for treating or preventing a
XX pathology associated with IFN polypeptide in a human. They are useful for
XX determining the presence of or predisposition to a disease associated
XX with altered levels of IFN polypeptide or polynucleotide. IFN nucleic
XX acids, polypeptides and antibodies are useful for diagnosis, prevention
XX or treatment of variety of immunological and cell proliferative
XX disorders, such as autoimmune diseases e.g. lupus erythematosus,
XX immunodeficiency diseases such as acquired immunodeficiency syndrome
XX (AIDS), graft rejection, viral infections including hepatitis and human
XX immunodeficiency virus (HIV), immune-mediated glomerulonephritis,
XX haematologic diseases such as aplastic anaemia and chronic neutropenia
XX and cancer. In addition they are also useful for treating or diagnosing
XX various disorders associated with cell death, including myocardial
XX infarction, stroke, neurologic diseases including Alzheimer's and
XX Parkinson's disease, amyotrophic lateral sclerosis and spinal muscular
XX atrophy. IFN nucleic acids and polypeptides are also useful for
XX identifying interferon-like proteins and interferon agonists, for
XX screening drugs and compounds which inhibit or enhance IFN activity or
XX function and as targets for the identification of small molecules that
XX are immunostimulatory, immunosuppressive, or stimulate or suppress normal
XX cell or tumour cell growth in mammals, including humans.
XX
XX Sequence 1270 BP; 401 A; 271 C; 279 G; 319 T; 0 other;
XX
XX Query Match 23.9%; Score 803; DB 22; Length 1270;
XX Best Local Similarity 95.5%; Pred. No. 9,2e-189;
XX Matches 853; Conservative 0; Mismatches 0; Indels 4; Gaps 1;
QY 45 GCACAGAGGGGTAGACCTGCTTCTCTTAAGTGGAGCGGACAGGCGGAGCAGATTCACA 104
    |||||||
DB 67 GCACAGAGGGGTAGACCTGCTTCTCTTAAGTGGAGCGGAGCGGAGCAGCAGATTCACA 126
QY 105 CCGTCCCGGACAGCAACAGACCATCTGCTGGGAGAACCCCTGCTCTCTTAAGAG 164
    |||||||
DB 127 CCGTCCCGGACAGCAACAGACCATCTGCTGGGAGAACCCCTGCTCTCTCTCTCTTA 186
QY 165 AAGATGTGCAATGGGTATTCACAGAGAGCAATTTCCGCTATCTCATCTGCAATTCAG 224
    |||||||
DB 187 AAGATGTGCAATGGGTATTCACAGAGAGCAATTTCCGCTATCTCATCTGCAATTCAG 246
QY 225 GGCACAGGTGAAATGTACATCCAGGTGGAGCCTGTGCTGGACTACCTGCTTTCTGAC 284
    |||||||
DB 247 GGCACAGGTGAAATGTACATC----- 268

```


QY 582 CATGAGAGAGAACTGTTGACATTTGAACAGAAACCGGATTGCTCTGAGAAACAA 641
 DB 567 CATGAGAGAGAACTGTTGACATTTGAACAGAAACCGGATTGCTCTGAGAAACAA 626
 QY 642 TGGAAATGATGAGTGTAGAGAGCTACTAAAGAGATTGTGAGAAAGAA-CTGGT 700
 DB 627 TGGAAATGATGAGTGTAGAGAGCTACTAAAGAGATTGTGAGAAAGAA-CTGGT 686
 QY 701 TCTGTCATTTGATGATGTTCTTCTGCAACAGGAA-CAATGAACTTCTCAAGCTTA 759
 DB 687 TCTGTCATTTGATGATGTTCTTCTGCAACAGGAA-CAATGAACTTCTCAAGCTTA 746
 QY 760 ACAGGCTGATTTGCTGCAAGAAAGATGAGATTGATTTGCAATTTGATGCT 819
 DB 747 ACAGGCTGATTTGCTGCAAGAAAGATGAGATTGATTTGCAATTTGATGCT 806
 QY 820 CCTCAAGTGAAGAGCACTTCTTCAACAGGATTCAGCAAACTGAGAGAGGTC 879
 DB 807 CCTCAAGTGAAGAGCACTTCTTCAACAGGATTCAGCAAACTGAGAGAGGTC 866
 QY 880 TGGGGATGAGATTAATCATCATCATCATCTTTTGCAGATTTCTCTGATTTGAG 937
 DB 867 TGGGGATGAGATTAATCATCATCATCATCTTTTGCAGATTTCTCTGATTTGAG 924

RESULT 13

AA01150
 ID AA01150 standard: DNA: 3692 BP.

AC AA01150:

DT 12-SEP-2001 (first entry)

DE Interferon induced nucleic acid, IFN5.

KW Interferon induced nucleic acid; autoimmune disease; lupus erythematosus;

KW immunodeficiency; stroke; acquired immunodeficiency syndrome; AIDS;

KW graft rejection; viral infection; hepatitis; aplastic anemia; cancer;

KW human immunodeficiency virus; HIV; immune-mediated glomerulonephritis;

KW haematologic disease; chronic neutropenia; myocardial infarction;

KW neurological disease; Alzheimer's disease; Parkinson's disease; tumour;

KW amyotrophic lateral sclerosis; spinal muscular atrophy; human; IFN5; ds.

KW XX

OS Homo sapiens.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Location/Qualifiers
 1..804
 /tag- a
 /product- "IFN5"
 /note- "Interferon induced polypeptide"

MO200118208-A2.

15-MAR-2001.

08-SEP-2000; 2000MO-US24704.

08-SEP-1999; 9905-0152921.

20-OCT-1999; 9905-0160575.

20-JAN-2000; 2000US-0177104.

07-SEP-2000; 2000US-0656633.

XX

XX

XX

XX

XX

XX

New interferon induced polypeptides and polynucleotides useful for the diagnosis, prevention and treatment of immunological, all

PT proliferative disorders, such as lupus erythematosus, cancer, stroke
 PT and Alzheimer's disease
 XX
 PS Claim 9: Page 33-39; 134pp: English.

XX The sequence represents interferon induced nucleic acid, IFN5. IFN
 CC nucleic acids and polypeptides are useful for treating or preventing a
 CC pathology associated with IFN polypeptide in a human. They are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of IFN polypeptide or polynucleotide. IFN nucleic
 CC acids, polypeptides and antibodies are useful for diagnosis, prevention
 CC or treatment of variety of immunological and cell proliferative
 CC disorders, such as autoimmune diseases e.g. lupus erythematosus,
 CC immunodeficiency diseases such as acquired immunodeficiency syndrome
 CC (AIDS), graft rejection, viral infections including hepatitis and human
 CC immunodeficiency virus (HIV), immune-mediated glomerulonephritis,
 CC haematologic diseases such as aplastic anaemia and chronic neutropenia
 CC and cancer. In addition they are also useful for treating myocardial
 CC various disorders associated with cell death, including myocardial
 CC infarction, stroke, neurological diseases including Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis and spinal muscular
 CC atrophy. IFN nucleic acids and polypeptides are also useful for
 CC identifying interferon-like proteins and interferon agonists, for
 CC screening drugs and compounds which inhibit or enhance IFN activity or
 CC function and as targets for the identification of small molecules that
 CC are immunostimulatory, immunosuppressive, or stimulate or suppress normal
 CC cell or tumour cell growth in mammals, including humans.

Sequence 3692 BP; 1166 A; 709 C; 806 G; 1011 T; 0 other:

Query Match 19.7%; Score 663; DB 22; Length 3692;

Best Local Similarity 92.9%; Pred. No. 8.2e-154; Matches 722; Conservative 0; Mismatches 15; Indels 40; Gaps 1;

QY 45 GCACAGAGCGTAGACCTCTCTTAAGTGGCAGCGGACGCGCAGCAATTTCA 104
 DB 55 GCACAGAGCGTAGACCTCTCTTAAGTGGCAGCGGACGCGCAGCAATTTCA 114
 QY 105 CCTGTCGCGGACAGCAAGCAACCATCTCTGGGAAACCTCTCTCTCTGAGAAAG 164
 DB 115 CCTGTCGCGGACAGCAAGCAACCATCTCTGGGAAACCTCTCTCTCTGAGAAAG 174
 QY 165 AAGATGTCGATGGGATTTCCACAGAGAGAAATTTCCGATTCATCTCTT 224
 DB 175 AAGATGTCGATGGGATTTCCACAGAGAGAAATTTCCGATTCATCTCTT 234
 QY 225 GGCACAGGTTGAATGTACATTCAGAGTGGAGCTGTGCTGACATTTCTGCC 284
 DB 235 GGCACAGGTTGAATGTACATTCAGAGTGGAGCTGTGCTGACATTTCTGCC 256
 QY 285 TGCAGAGGTGAAGAGCAGATTCACAGAGCACTGCCACCTCCGGAAACATTAAGT 344
 DB 257 --CAGAGGTGAAGAGCAGATTCACAGAGCACTGCCACCTCCGGAAACATTAAGT 314
 QY 345 TGAATGCTGCTGACACCTTGGAGAGGAGTGTGGACCTTGTGAGCTGTAAT 404
 DB 315 TGAATGCTGCTGACACCTTGGAGAGGAGTGTGGACCTTGTGAGCTGTAAT 374
 QY 405 CGTGGAGGCTCTCGGAGAACCGGACGCTTGGCCGCGCGCTACATGAATGTGACCT 464
 DB 375 CGTGGAGGCTCTCGGAGAACCGGACGCTTGGCCGCGCGCTACATGAATGTGACCT 434
 QY 465 CACGACTTGCCTCTCCATCTGTTGAGAACGCTATATGATATCTCCAACTTCTTA 524
 DB 435 CACGACTTGCCTCTCCATCTGTTGAGAACGCTATATGATATCTCCAACTTCTTA 494
 QY 525 CCTCTTACGCTGCTGAGACCTTGGAGAGCTTGTAGAGAGCTTGTAGAAATGAT 584
 DB 495 CCTCTTACGCTGCTGAGACCTTGGAGAGCTTGTAGAGAGCTTGTAGAAATGAT 554
 QY 585 GGAGAGGAACCTGTGACATTTGAAGAGAGAAACCGGATGTGCTGAGAGAAATATGG 644
 DB 555 GGAGAGGAACCTGTGACATTTGAAGAGAGAAACCGGATGTGCTGAGAGAAATATGG 614

OY 645 AATGATCAGGTGTAGAGAGCTACTAAAAAGATTGTGAGAAAGAAATTTGGTTTC 704
 DB 615 AATGATCAGGTGTAGAGAGCTACTAAAAAGATTGTGAGAAAGAAATTTGGTTTC 674
 OY 705 TGCATTTCGATGTTCTTCGTCACAAAGCAATGACCTTGTCCAAATTAACAGG 764
 DB 675 TGCATTTCGATGTTCTTCGTCACAAAGCAATGACCTTGTCCAAATTAACAGG 734
 OY 765 CTCTGATTGCTGAGAAAGCAATGACAGATTGATGATTCACAAAGTTATGGTCC 821
 DB 735 CTCTGATTGCTGAGAAAGCAATGACAGATTGATGATTCACAAAGTTATGGTCC 791
 RESULT 14
 AAS01152
 ID AAS01152 standard; DNA; 3704 BP.
 AC AAS01152;
 DT 12-SEP-2001 (first entry)
 DE Interferon induced nucleic acid, IFN7.
 KM Interferon induced nucleic acid; autoimmune disease; lupus erythematosus;
 KM immunodeficiency; stroke; acquired immunodeficiency syndrome; AIDS;
 KM graft rejection; viral infection; hepatitis; aplastic leukemia; cancer;
 KM human immunodeficiency virus; HIV; immune-mediated glomerulonephritis;
 KM haematologic disease; chronic neutropenia; myocardial infarction;
 KM neurological disease; Alzheimer's disease; Parkinson's disease; tumour;
 KM amyotrophic lateral sclerosis; spinal muscular atrophy; human; IFN7; ds.
 OS Homo sapiens.
 FT CDS 1..816 Location/Qualifiers
 FT /tag= "IFN7"
 FT /product= "Interferon induced poly(peptide)"
 FT /note= "Interferon induced poly(peptide)"
 PN MO200118208-A2.
 PD 15-MAR-2001.
 PF 08-SEP-2000; 2000MO-US24704.
 PR 08-SEP-1999; 99US-0152921.
 PR 20-OCT-1999; 99US-0160575.
 PR 20-JAN-2000; 2000US-0177104.
 PR 07-SEP-2000; 2000US-0656633.
 PA (CURA-) CURAGEN CORP.
 PA (BIOJ) BIOGEN INC.
 PI Peyman JA, Da Silva A, Hochman P, Hsu A;
 DR WPI: 2001-235201/24.
 DR P-PSDB: AA000299.
 PT New interferon induced polypeptides and polynucleotides, useful for the
 PT diagnosis, prevention and treatment of immunological, cell
 PT proliferative disorders, such as lupus erythematosus, cancer, stroke
 PT and Alzheimer's disease
 PS Claim 9; Page 44-50; 134pp; English.
 CC The sequence represents interferon induced nucleic acid, IFN7. IFN
 CC nucleic acids and polypeptides are useful for treating or preventing a
 CC pathology associated with IFN polypeptide in a human. They are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of IFN polypeptide or polynucleotide. IFN nucleic
 CC acids, polypeptides and antibodies are useful for diagnosis, prevention
 CC or treatment of variety of immunological and cell proliferative

CC disorders, such as autoimmune diseases e.g. lupus erythematosus,
 CC immunodeficiency diseases such as acquired immunodeficiency syndrome
 CC (AIDS), graft rejection, viral infections including hepatitis and human
 CC immunodeficiency virus (HIV), immune-mediated glomerulonephritis,
 CC haematologic diseases such as aplastic anaemia and chronic neutropenia
 CC and cancer. In addition they are also useful for treating or diagnosing
 CC various disorders associated with cell death, including myocardial
 CC infarction, stroke, neurological diseases including Alzheimer's and
 CC Parkinson's diseases, amyotrophic lateral sclerosis and spinal muscular
 CC atrophy. IFN nucleic acids and polypeptides are also useful for
 CC identifying interferon-like proteins and interferon agonists, for
 CC screening drugs and compounds which inhibit or enhance IFN activity or
 CC function and as targets for the identification of small molecules that
 CC are immunostimulatory, immunosuppressive, or stimulate or suppress normal
 CC cell or tumour cell growth in mammals, including humans.
 XX Sequence 3704 BP; 1168 A; 713 C; 809 G; 1014 T; 0 other;
 SO
 Query Match 19.7%; Score 663; DB 22; Length 3704;
 Best Local Similarity 92.9%; Pred. No. 8.2e-154;
 Matches 722; Conservative 0; Mismatches 15; Indels 40; Gaps 1;
 OY 45 GCACAGAGCGGTAGACCCCTGCTCTCTAGTGGGAGCGGACAGGCGACCAATTTCA 104
 DB 67 GCACAGAGCGGTAGACCCCTGCTCTCTAGTGGGAGCGGACAGGCGACCAATTTCA 126
 OY 105 CCTGTCCCGGACAGCAACAGCAACATCTGCTTGGAGAACCCCTCTCCCTTCTTA 164
 DB 127 CCTGTCCCGGACAGCAACAGCAACATCTGCTTGGAGAACCCCTCTCCCTTCTTA 186
 OY 165 AAGATGTGGAATGGGTATTCACAGACGAGCAATTTCCGTATCTCATCTCTCTG 224
 DB 187 AAGATGTGGAATGGGTATTCACAGACGAGCAATTTCCGTATCTCATCTCTCTG 246
 OY 225 GGGCCAGGTTGAAATGTATCTACAGAGTGGAGCCCTGTGTGACATCTGCTG 284
 DB 247 GGGCCAGGTTGAAATGTATCTAC----- 268
 OY 285 TGCAGAGGTGAAGAGAGCAATTCAGAGAGACAGTGGCCACTCGGGGAACATTA 344
 DB 269 --CAGAGGTGAAGAGAGCAATTCAGAGAGACAGTGGCCACTCGGGGAACATTA 326
 OY 345 TGAATGCTGCTGAGACACCTTGGAGAAAGGAGTGTGGACCTTGGTGGACATTA 404
 DB 327 TGAATGCTGCTGAGACACCTTGGAGAAAGGAGTGTGGACCTTGGTGGACATTA 386
 OY 405 COTGAGAGCCCTCCGAGAGACGGGAGCCCTCTGACCCGCTACATGAATTAAGCT 464
 DB 387 COTGAGAGCCCTCCGAGAGACGGGAGCCCTCTGACCCGCTACATGAATTAAGCT 446
 OY 465 CAGGACTTGGCCCTCTGATCGTTTGAAGAGCGCTCATGATGATATATCTCCAA 524
 DB 447 CAGGACTTGGCCCTCTGATCGTTTGAAGAGCGCTCATGATGATATATCTCCAA 506
 OY 525 CCTCTTTAGCCCACTCTGCTGGAGCAACCTTCTAGTATGAACGCTTTGGATAAT 584
 DB 507 CCTCTTTAGCCCACTCTGCTGGAGCAACCTTCTAGTATGAACGCTTTGGATAAT 566
 OY 585 GGAGAGGAACCTGTTGACATTTGAAGACAGAACCGGATTCTGCTGAGAAATTAAG 644
 DB 567 GGAGAGGAACCTGTTGACATTTGAAGACAGAACCGGATTCTGCTGAGAAATTAAG 626
 OY 645 AATGATCAGGTGTAGAGAGCTACTAAAAAGATTGTGAGAAAGAAATTTGGTTTC 704
 DB 627 AATGATCAGGTGTAGAGAGCTACTAAAAAGATTGTGAGAAAGAAATTTGGTTTC 686
 OY 705 TGCATTTCGATGTTCTTCGTCACAAAGCAATGACCTTGTCCAAATTAACAGG 764
 DB 687 TGCATTTCGATGTTCTTCGTCACAAAGCAATGACCTTGTCCAAATTAACAGG 746
 OY 765 CTCTGATTGCTGAGAAAGCAATGACAGATTGATGATTCACAAAGTTATGGTCC 821
 DB 747 CTCTGATTGCTGAGAAAGCAATGACAGATTGATGATTCACAAAGTTATGGTCC 803

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

OM nucleic - nucleic search, using sw model

Run on: May 16, 2003, 12:54:36 ; Search time 8627 seconds

11351.680 Million cell updates/sec

Title: US-09-515-363C-1

Sequence: 1 gcgcgcgcgcctgagagccc.....aaatgatgtt | taccctc 3365

Scoring table: IDENTITY_NUC

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 410+.30

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 08

Maximum Match 100%
Listing first 45 summaries

Database :

1:	gb_ba:*
2:	gb_hlg:*
3:	gb_in:*
4:	gb_om:*
5:	gb_ov:*
6:	gb_pat:*
7:	gb_ph:*
8:	gb_pl:*
9:	gb_pr:*
10:	gb_ro:*
11:	gb_sts:*
12:	gb_sy:*
13:	gb_un:*
14:	gb_vl:*
15:	em_ba:*
16:	em_fun:*
17:	em_hum:*
18:	em_in:*
19:	em_mu:*
20:	em_om:*
21:	em_ov:*
22:	em_ov:*
23:	em_ph:*
24:	em_ph:*
25:	em_pl:*
26:	em_ro:*
27:	em_sts:*
28:	em_un:*
29:	em_vl:*
30:	em_hlg_hum:*
31:	em_hlg_in:*
32:	em_hlg_other:*
33:	em_hlg_mus:*
34:	em_hlg_pln:*
35:	em_hlg_rtd:*
36:	em_hlg_nam:*
37:	em_hlg_vlt:*
38:	em_sy:*
39:	em_hlg_hum:*
40:	em_hlg_mus:*
41:	em_hlg_other:*

Pred. No. is the number of results predicted by chance to have a

Result No.	Score	Query Match	length	DB	ID	Description
1	3365	100.0	3380	9	AF095844	AF095844 Homo sapi
2	3338.8	99.2	3372	6	AX300832	AX300832 Sequence
3	3338.8	99.2	3373	6	AY017378	AY017378 Homo sapi
4	2202.4	65.5	3771	10	AF374384	AF374384 Mus musc
5	2135.4	63.5	3078	10	AY075132	AY075132 Mus musc
6	1766	52.5	1776	9	AK056293	AK056293 Homo sapi
7	1435.6	42.7	1443	6	AX300841	AX300841 Sequence
8	1402	41.7	2468	6	BC025508	BC025508 Mus musc
9	1254.6	37.3	1284	6	AX300838	AX300838 Sequence
10	1194.4	35.5	1213	9	BC007066	BC007066 Homo sapi
11	1186	35.2	2050	10	BC004031	BC004031 Mus musc
12	803	23.9	1258	6	AX098232	AX098232 Sequence
13	803	23.9	1270	6	AX098236	AX098236 Sequence
14	663	19.7	3692	6	AX098234	AX098234 Sequence
15	663	19.7	3704	6	AX098238	AX098238 Sequence
16	623.2	18.5	95417	9	AC010876	AC010876 Homo sapi
17	384	11.4	923	10	BC018605	BC018605 Mus musc
18	371.2	11.0	392	6	AX071769	AX071769 Sequence
19	356.8	10.6	387	6	AX071783	AX071783 Sequence
20	303	9.0	163681	9	AC007750	AC007750 Homo sapi
21	285.4	8.5	301	6	150897	150897 Sequence
22	284.4	8.5	118931	6	AC094719	AC094719 Rattus no
23	284.4	8.5	182906	2	AC011919	AC011919 Rattus no
24	279.4	8.3	210317	2	AC115074	AC115074 Mus musc
25	269.2	8.0	377	6	A74554	A74554 Sequence 24
26	269.2	8.0	377	6	A77533	A77533 Sequence 24
27	223.4	6.6	2326	10	AF316999	AF316999 Mus musc
28	221.8	6.6	2336	10	BC029209	BC029209 Mus musc
29	179.6	5.3	2584	9	BC014949	BC014949 Homo sapi
30	178	5.3	2613	9	AK021416	AK021416 Homo sapi
31	172	5.1	182906	2	AC111919	AC111919 Rattus no
32	166.8	5.0	118891	2	AC090419	AC090419 Rattus no
33	153	4.5	164	6	AX300834	AX300834 Sequence
34	150.2	4.5	2250	9	AK097669	AK097669 Homo sapi
35	148.6	4.4	3065	9	AF038963	AF038963 Homo sapi
36	141.8	4.2	160	6	A74468	A74468 Sequence 15
37	141.8	4.2	160	6	A77447	A77447 Sequence 15
38	130	3.9	112361	9	AC027316	AC027316 Homo sapi
39	130	3.9	143200	9	AC008413	AC008413 Homo sapi
40	124.6	3.7	193	6	150896	150896 Sequence 6
41	114	3.4	2623	4	AF181119	AF181119 Sus scrof
42	109.6	3.3	4034	4	HS480236	HS480236 Homo sapi
43	109.6	3.3	3076	9	AK023661	AK023661 Homo sapi
44	72.4	2.2	100635	9	AC003104	AC003104 Homo sapi
45	72.4	2.2	154808	2	AC105024	AC105024 Homo sapi

ALIGNMENTS

RESULT 1	
AF095844	
LOCUS	3380 bp mRNA linear ref U73623
DEFINITION	Homo sapiens melanoma differentiation associated protein-5 (MDA5)
ACCESSION	AF095844
VERSION	AF095844.1 GI:11344593
KEYWORDS	
SOURCE	Homo sapiens.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 3380)
AUTHORS	Kang,D.C., Gopalakrishnan,R.V., Wu,Q., Jankowsky,E., Pye,A.M. and Fisher,P.B.

QY 1441 GAAATGAGAGAGTCTGGTTCATATGTCAGACTTCCCTCATTAATTGATGAA 1500
 DB 1441 GAAATGAGAGAGTCTGGTTCATATGTCAGACTTCCCTCATTAATTGATGAA 1500
 QY 1501 TGTATCAGACCAAGAGAGAGTGTATATACATATGAGGCAATTTGATGCAAG 1560
 DB 1501 TGTATCAGACCAAGAGAGAGTGTATATACATATGAGGCAATTTGATGCAAG 1560
 QY 1561 AACTGAAAACATAGACTGAGAAAAGAAACAAACAGTGTATGCTGCTGCTGAGATA 1620
 DB 1561 AACTGAAAACATAGACTGAGAAAAGAAACAAACAGTGTATGCTGCTGCTGAGATA 1620
 QY 1621 CTGGAGCAACAGCTTCACTGGTGTGGAGGGGCCAGAGCAAGCAAGCAAGTGAAGAA 1680
 DB 1621 CTGGAGCAACAGCTTCACTGGTGTGGAGGGGCCAGAGCAAGCAAGCAAGTGAAGAA 1680
 QY 1681 CACATTTTAAACTATGTCGCAATCTTGATGCACTTTACTTAAACTGTAAGAAGAAC 1740
 DB 1681 CACATTTTAAACTATGTCGCAATCTTGATGCACTTTACTTAAACTGTAAGAAGAAC 1740
 QY 1741 CTGTATCACTGAAAAACCAATACAGAGCCATGCAACAGTGTGCAATGAGATGCA 1800
 DB 1741 CTGTATCACTGAAAAACCAATACAGAGCCATGCAACAGTGTGCAATGAGATGCA 1800
 QY 1801 ACCAGAGAGATCCATTTTAAAGAAACTCTAGAAATATGACAGAGATTAACCTTAT 1860
 DB 1801 ACCAGAGAGATCCATTTTAAAGAAACTCTAGAAATATGACAGAGATTAACCTTAT 1860
 QY 1861 TGTCAATGAGTCCATGTCAGATTTTGGAACTCAACCTTATGCAATGAGATGCA 1920
 DB 1861 TGTCAATGAGTCCATGTCAGATTTTGGAACTCAACCTTATGCAATGAGATGCA 1920
 QY 1921 ATGGAAGAAAAAAGCTGCAAAAAAGAAATCCGAAAGAAAGTGTGTGTAAGCAATTG 1980
 DB 1921 ATGGAAGAAAAAAGCTGCAAAAAAGAAATCCGAAAGAAAGTGTGTGTAAGCAATTG 1980
 QY 1981 AGGAGATACATAGAGCCCTACAAATTAATGACACATGGAATGATATAGTATAGT 2040
 DB 1981 AGGAGATACATAGAGCCCTACAAATTAATGACACATGGAATGATATAGTATAGT 2040
 QY 2041 CATCTTGAACCTTTCTATATGAGAGAAAGATAGAGTTTCCAGTCAATAGATGAT 2100
 DB 2041 CATCTTGAACCTTTCTATATGAGAGAAAGATAGAGTTTCCAGTCAATAGATGAT 2100
 QY 2101 AGTATGAGGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2160
 DB 2101 AGTATGAGGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2160
 QY 2161 AACTTTGAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2220
 DB 2161 AACTTTGAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2220
 QY 2221 AAAATGTTGAAAAGCTGCTGAAAACCCAGAAATATGAAATGAAAGTGTGCAATTA 2280
 DB 2221 AAAATGTTGAAAAGCTGCTGAAAACCCAGAAATATGAAATGAAAGTGTGCAATTA 2280
 QY 2281 AGAATATACATATGAGCAATATGATGAGAGTGAAGGAAAGAGAGAGATATATAT 2340
 DB 2281 AGAATATACATATGAGCAATATGATGAGAGTGAAGGAAAGAGAGAGATATATAT 2340
 QY 2341 ACAAAAACAGAGAGATGATGCTTCCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 2400
 DB 2341 ACAAAAACAGAGAGATGATGCTTCCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTG 2400
 QY 2401 GCTGAAGTAGAGTCAAAAGCCACCATCTGATTTGAGAGTGAAGAGAGAGAGAGAG 2460
 DB 2401 GCTGAAGTAGAGTCAAAAGCCACCATCTGATTTGAGAGTGAAGAGAGAGAGAGAG 2460
 QY 2461 CCATGACAG 2520
 DB 2461 CCATGACAG 2520
 QY 2521 CTGCTATGCTTACACAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2580

DB 2521 CTGCTATGCTTACACAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2580
 QY 2581 ATCCGTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2640
 DB 2581 ATCCGTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2640
 QY 2641 GCTGATGAG 2700
 DB 2641 GCTGATGAG 2700
 QY 2701 ACAGTTATGATTTCCGAG 2760
 DB 2701 ACAGTTATGATTTCCGAG 2760
 QY 2761 AAACAG 2820
 DB 2761 AAACAG 2820
 QY 2821 AAATGAAAACAG 2880
 DB 2821 AAATGAAAACAG 2880
 QY 2881 TTCTTTGCAAAAAGCTGAGTGTGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2940
 DB 2881 TTCTTTGCAAAAAGCTGAGTGTGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2940
 QY 2941 AAATGATCAGTCAATATGACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3000
 DB 2941 AAATGATCAGTCAATATGACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3000
 QY 3001 AAAGCAGTCAAG 3060
 DB 3001 AAAGCAGTCAAG 3060
 QY 3061 GGCCAGAGCTTGGGAG 3120
 DB 3061 GGCCAGAGCTTGGGAG 3120
 QY 3121 AGGATTTTGTAGTGTGTTTCAAAAATTAATTAACAAAGAGAGAGAGAGAGAG 3180
 DB 3121 AGGATTTTGTAGTGTGTTTCAAAAATTAATTAACAAAGAGAGAGAGAGAGAG 3180
 QY 3181 GAATTAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3240
 DB 3181 GAATTAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3240
 QY 3241 GATTAGCAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3300
 DB 3241 GATTAGCAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3300
 QY 3301 TGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3360
 DB 3301 TGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3360
 QY 3361 CTCTG 3365
 DB 3361 CTCTG 3365
 RESULT 2
 AX300832 3372 bp DNA linear SAT 30-NOV-2001
 LOCUS AX300832
 DEFINITION Sequence 1 from Patent WO0185955.
 ACCESSION AX300832
 VERSION AX300832.1 GI:17382110
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1
 Bahr, G., Cocude, C. and Capron, A.

TITLE rh16 polypeptides and its fragments and polynucleotides encoding
 JOURNAL said polypeptides and its therapeutic uses
 Patent: WO 0185955-A1 15-NOV-2001;
 Istac (FR) : INSTITUT PASTEUR DE LILIE (FR)
 FEATURES Location/Qualifiers
 source 1.3372
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 155.3332
 /note="unnamed protein product"
 /protein_id="CAD13049.1"
 /codon_start=1
 /db_xref="gi:17382111"
 /translation="MSGSTDENRFLVISCFRANVKKHVEPYLYLFLPAEVE
 QIQRTVATSGNMAVELLSTLEKVMHGMREVEAIPRTGSLARVYNPELIDL
 PSPFNADHEYLQNLLOPLVDKLVIRVLDLQCMELTITIEDNRHIAAENGNG
 EGVRELLKRIYQENMFSAFLNRLQNGELVDELTPNCGSEGATLQAEHLIDP
 OVEBOLLSTTQPNLEKYMENSSSPADSVVSVSLTSLAEQSVSLDESLGH
 NSNMGSDSGTMSDSDSENVAAKASPRELQRLPYQMEVAPALRGKNIITCLPTGSG
 KTRAVVITAKDLKKRASEPKVIVLVNVLVEQLVPEEPFLKMYRVLGLSG
 DTOLKISPEPVVSCDIIISTQLLENSLNLNEDNAVOLDSFPIIDECHTNK
 EAVNNINRHYLMOKLNNRLKKNKPVIPYQLGLTASGVGATQAAEHLIDP
 LCAINDAFTIKVLENLDLQKMOIOEPCKKFAIDATPFIPEKELIPTRIQYCG
 MSPMDPTQPEQMAIOMEKKAKEGRKRSVCAKHLVINKALQINDTIRMDIAYT
 HLETPYNEKOKKRAVIDEDSDSDSDYCCODEDEDDKPKLIDETRLMTLTFE
 NKKMLKRLAENPEIENEKLTIRNTIMQVTRTESARLIPKTRISAVASQWITE
 NEKFAVEGVKAHLIGASHSEFPKPTQNEQEVLSKPIKINILIAVTAEBLIDP
 KECNTIVIRGLATNEIAYQARBARADESTVIVASHSVIEREINDEKEMMYK
 AHCQVNNKPREYAKHLELOMOSIMKMKTKRMAKRYNNPILITELCKNSVLA
 CSGEDIHVLEKHHYNNPREFELYIVENNALOKCAVNLNGEIPKCCQAACTMM
 VHKGLDLCIKRNVVYFKNNSTKQYKKVVELPITEINDISECLSPED"

BASE COUNT 1157 a 637 c 748 g 830 t
 ORIGIN

Query Match 99.2%; Score 3338.8; DB 6; Length 3372;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 3343; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

15 GAGCCTGTGGACACCTGTCATTTGTGAGCAGAGAGCGGTAGACCTCTTCTGTAG 74
 1 GGGCCTGTGGACACCTGTCATTTGTGAGCAGAGAGCGGTAGACCTCTTCTGTAG 60
 75 TGGGCGAGGAGACGCGGACGACGACATTTACCTGTCCCGACAGACAGATTCATCTGT 134
 61 TGGGCGAGGAGACGCGGACGACGACATTTACCTGTCCCGACAGACAGATTCATCTGT 120
 135 TGGGAGAACCTCTCTCCCTCTCTGAGAGAGAGAGATGTGAATGGGTATCTACAGAGA 194
 121 TGGGAGAACCTCTCTCCCTCTCTGAGAGAGAGAGATGTGAATGGGTATCTACAGAGA 180
 195 GAATTCGGCTATCTCATCTGCTTTCAGGGCGAGGAGTGAATAATGTATCTCAGGTGA 254
 181 GAATTCGGCTATCTCATCTGCTTTCAGGGCGAGGAGTGAATAATGTATCTCAGGTGA 240
 255 GCCTGTGTGAGTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 314
 241 GCCTGTGTGAGTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
 315 AGTGGCAGCTCCGCGAGACATGAGGAGCTTGAAGTGCCTGCTGAGACCTCTGAGAGAGG 374
 301 AGTGGCAGCTCCGCGAGACATGAGGAGCTTGAAGTGCCTGCTGAGACCTCTGAGAGAGG 360
 375 AGTGTGGACACTTGTGTGAGTGTGGAGATTCGTGAGAGATTCGTGAGAGATTCGTGAG 434
 361 AGTGTGGACACTTGTGTGAGTGTGGAGATTCGTGAGAGATTCGTGAGAGATTCGTGAG 420
 435 TGTGGCGCGCGCTACATGAACCTGAGTGCAGGAGCTTGGCTCTCCATCTTTTGAGAA 494
 421 TGTGGCGCGCGCTACATGAACCTGAGTGCAGGAGCTTGGCTCTCCATCTTTTGAGAA 480
 495 CGCTCATGATGATATCTGCAACTGCTGAACCTCTGAGGCGACATCTGAGACAGGT 554
 481 CGCTCATGATGATATCTGCAACTGCTGAACCTCTGAGGCGACATCTGAGACAGGT 540

555 TCTAGTAGAGACGCTTGTGATTAAGTGCATGAGAGAGAGACTGTGACAAATGAAACAG 614
 541 TCTAGTAGAGACGCTTGTGATTAAGTGCATGAGAGAGAGACTGTGACAAATGAAACAG 600
 615 AAACGGATTTGCTGCTGCAAGAAACAAATGGAATTAATCAGTGTAGAGAGCTTTAA 674
 601 AAACGGATTTGCTGCTGCAAGAAACAAATGGAATTAATCAGTGTAGAGAGCTTTAA 660
 675 AAGGATTTGCAAGAAAGAAATGAGTGTCTGCAATTTGGAATGTTCTGCTCAACAGG 734
 661 AAGGATTTGCAAGAAAGAAATGAGTGTCTGCAATTTGGAATGTTCTGCTCAACAGG 720
 735 AAACATGAACTTTGCAAGAGTTAAGAGCTGTGATTTGCTGAGAAAGCAATGAGAGAT 794
 721 AAACATGAACTTTGCAAGAGTTAAGAGCTGTGATTTGCTGAGAAAGCAATGAGAGAT 780
 795 TGAGAAATTTACAAAGTTGATAGGTGCTGCAAGTGTGAAGAGCAACTTTCTTAACAC 854
 781 TGAGAAATTTACAAAGTTGATAGGTGCTGCAAGTGTGAAGAGCAACTTTCTTAACAC 840
 855 TCAGCAAACTGAGAGAGAGTGTGGGCGATGAGAAATTAAGTATCATCATCATCTT 914
 841 TCAGCAAACTGAGAGAGAGTGTGGGCGATGAGAAATTAAGTATCATCATCATCTT 900
 915 TCAGAAATTTCTGATGTTTCAAGATTCAGACAGCAAGTTTGCAGAGAGAGTGTAGCTG 974
 901 TCAGAAATTTCTGATGTTTCAAGATTCAGACAGCAAGTTTGCAGAGAGAGTGTAGCTG 960
 975 CTTAATGAAAGTCTTGTGACATTAACAGCAATGAGGCGATGATTCAGGACCATTTGAG 1034
 961 CTTAATGAAAGTCTTGTGACATTAACAGCAATGAGGCGATGATTCAGGACCATTTGAG 1020
 1035 TGATTCAGATGAGAGAAATGTGGCAGCAAGAGCAATCCCGAGCGAGAACTGCAATGAG 1094
 1021 TGATTCAGATGAGAGAAATGTGGCAGCAAGAGCAATCCCGAGCGAGAACTGCAATGAG 1080
 1095 GCCTTACCAATGAGAGTGTGCCAGCAGCCTTGGAGAGGAGAAATATCATCATTTGCT 1154
 1081 GCCTTACCAATGAGAGTGTGCCAGCAGCCTTGGAGAGGAGAAATATCATCATTTGCT 1140
 1155 CCCTACAGAGAGTGTGAGAGAGAGAGTGTGATTAATTCGCAAGATCATCATTAACAA 1214
 1141 CCCTACAGAGAGTGTGAGAGAGAGAGTGTGATTAATTCGCAAGATCATCATTAACAA 1200
 1215 GAGAGAAAGAGATGAGAGCTGTGAGAAAGTATAGTCTTTCATTAAGTATAGTATAGT 1274
 1201 GAGAGAAAGAGATGAGAGCTGTGAGAAAGTATAGTCTTTCATTAAGTATAGTATAGT 1260
 1275 TGAACAGCTTTCCGCAAGAGAGTTCACCAATTTTGAAGAAATGATGATGATTTGG 1334
 1261 TGAACAGCTTTCCGCAAGAGAGTTCACCAATTTTGAAGAAATGATGATGATTTGG 1320
 1335 ATTAAGTGTGATTAACCAAGTGAAGAAATATCATTTTCGCAAGATCATCATTAACAA 1394
 1321 ATTAAGTGTGATTAACCAAGTGAAGAAATATCATTTTCGCAAGATCATCATTAACAA 1380
 1395 TATTATCAGTACAGTCAATTCCTTGAAGAACTCCCTTTAAACTGTGAAGAAAGAGAA 1454
 1381 TATTATCAGTACAGTCAATTCCTTGAAGAACTCCCTTTAAACTGTGAAGAAAGAGAA 1440
 1455 TGTGTGTGTTCAATTTGAGAGCTTTTCCCTCAATTAATGATGATGATGATTAATTAAC 1514
 1441 TGTGTGTGTTCAATTTGAGAGCTTTTCCCTCAATTAATGATGATGATGATTAATTAAC 1500
 1515 CAAAGAGAGATGATTAATTAATCATCATGAGGATTTATGATGAGAGAGTGTGAAGAA 1574
 1501 CAAAGAGAGATGATTAATTAATCATCATGAGGATTTATGATGAGAGAGTGTGAAGAA 1560
 1575 TAGACTCAAGAGAAAGAAAGCAAGATGATTCCTTCCCTCAGATGATGAGAGTAAACAG 1634
 1561 TAGACTCAAGAGAAAGAAAGCAAGATGATTCCTTCCCTCAGATGATGAGAGTAAACAG 1620

1635 TTCACCTGCTGTGGAGGGGCCACGAGCAAGCAAGCCTTAAGCT 1694
 1621 TTCACCTGCTGTGGAGGGGCCACGAGCAAGCCTTAAGCT 1680
 1695 ATGTGCAATCTTGTGATGATTTACTATTAACTGTTAAGAAAGCTTA 1754
 1681 ATGTGCAATCTTGTGATGATTTACTATTAACTGTTAAGAAAGCTTA 1740
 1755 AAAACCAATACAGAGCCATGACAGAGCTTCCCATGACATGACACAC 1814
 1741 AAAACCAATACAGAGCCATGACAGAGCTTCCCATGACATGACACAC 1800
 1815 ATTTAAGAGAACTCTAGTAATATGACAGAGATTCACCTATTGTAATG 1874
 1801 ATTTAAGAGAACTCTAGTAATATGACAGAGATTCACCTATTGTAATG 1860
 1875 AATGTCAGATTTTGGAACTCAACCTATGACAGATGGGCAATTCAAAT 1934
 1861 AATGTCAGATTTTGGAACTCAACCTATGACAGATGGGCAATTCAAAT 1920
 1935 TCGAAAAAGAGAAATCGCAAGAAAGCTTTTGTGCAAGAACATTTGAG 1994
 1921 TCGAAAAAGAGAAATCGCAAGAAAGCTTTTGTGCAAGAACATTTGAG 1980
 1995 GGGCCCTACAAATTAATGACACATTTGATGATGATGATGATGATGAT 2054
 1981 GGGCCCTACAAATTAATGACACATTTGATGATGATGATGATGATGAT 2040
 2055 CTATATGAGAGAGAAAGATGAGAGATTTGACATGATGATGATGATGAT 2114
 2041 CTATATGAGAGAGAAAGATGAGAGATTTGACATGATGATGATGATGAT 2100
 2115 TGAATGATGATTTGATGATGATGATGATGATGATGATGATGATGATGAT 2174
 2101 TGAATGATGATTTGATGATGATGATGATGATGATGATGATGATGATGAT 2160
 2175 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2234
 2161 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2220
 2235 GCTGGCTGAAAAAGCAATATGAAAAATGAAAAAGCTGATTAATTAAT 2294
 2221 GCTGGCTGAAAAAGCAATATGAAAAATGAAAAAGCTGATTAATTAAT 2280
 2295 GAGGCAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2354
 2281 GAGGCAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2340
 2355 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2414
 2341 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2400
 2415 GAAAGCCCAATCTGATTTGAGAGCTGAGACAGAGCTGATGATGATGAT 2474
 2401 GAAAGCCCAATCTGATTTGAGAGCTGAGACAGAGCTGATGATGATGAT 2460
 2475 TGAAGCAAAAAAGAGTATTAATTAATTTGAGAGCTGAGACAGAGCTGAT 2534
 2461 TGAAGCAAAAAAGAGTATTAATTAATTTGAGAGCTGAGACAGAGCTGAT 2520
 2535 CACAGTGGAGAGAGAGTCTGATATTAAGAGATTAATGATGATGATGATGAT 2594
 2521 CACAGTGGAGAGAGAGTCTGATATTAAGAGATTAATGATGATGATGATGAT 2580
 2595 CGTACCAATGAATATGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCT 2654
 2581 CGTACCAATGAATATGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCT 2640
 2655 CTGAGCTGGTGTGCTGACAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCT 2714
 2641 CTGAGCTGGTGTGCTGACAGAGCTGAGAGCTGAGAGCTGAGAGCTGAGAGCT 2700
 2715 CCGAGAGAGATGATGATTAAGAGCTATGATGATGATGATGATGATGATGAT 2774

2701 CCGAGAGAGATGATGATTAAGAGCTATGATGATGATGATGATGATGATGAT 2760
 2775 TGCTCATAGATTTTGGATTAATGATGATGATGATGATGATGATGATGAT 2834
 2761 TGCTCATAGATTTTGGATTAATGATGATGATGATGATGATGATGATGAT 2820
 2835 GAGAAATATGCGAGATTAATGATGATGATGATGATGATGATGATGATGAT 2894
 2821 GAGAAATATGCGAGATTAATGATGATGATGATGATGATGATGATGATGAT 2880
 2895 CTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2954
 2881 CTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2940
 2955 CAATATGACCCAGATTAATGATGATGATGATGATGATGATGATGATGAT 3014
 2941 CAATATGACCCAGATTAATGATGATGATGATGATGATGATGATGATGAT 3000
 3015 GAGTGTGCGGATGATGATTAATGATGATGATGATGATGATGATGATGAT 3074
 3001 GAGTGTGCGGATGATGATTAATGATGATGATGATGATGATGATGATGAT 3060
 3075 AACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3134
 3061 AACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3120
 3135 GGTTCAAAAATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 3194
 3121 GGTTCAAAAATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 3180
 3195 ATTCCCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3254
 3181 ATTCCCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3240
 3255 TGAAGATTTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3314
 3241 TGAAGATTTTAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3300
 3315 ATTATGCTACAGACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 3364
 3301 ATTATGCTACAGACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 3350

RESULT 3
 LOCUS AY017378 3373 bp mRNA linear PRI 31-JAN-2001
 DEFINITION Homo sapiens RNA helicase-DEAD box protein RHL16 mRNA, complete cds.
 ACCESSION AY017378
 VERSION AY017378.1 GI:12621065
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens.
 Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 3373)
 Cocude,C., Kolesnichenko,V., Billaut-Mulot,O., Truong,M.-J., Capron,A. and Bahr,G.M.
 Identification of a new RNA helicase (RHL16) regulated by the immunomodulator Murabutide
 Unpublished
 JOURNAL 2 (bases 1 to 3373)
 Cocude,C., Kolesnichenko,V., Billaut-Mulot,O., Truong,M.-J., Capron,A. and Bahr,G.M.
 Direct Submission
 Submitted (09-JAN-2001) Laboratoire d'Immunologie Molculaire de l'Infection et de l'Inflammation, Institut Pasteur de Lille, 1 rue du Professeur Calmette BP 245, Lille 59019, France
 FEATURES
 source
 1..3373
 /organism="Homo sapiens"
 /db_xref="taxon:9606"

CDS	155..3232	/codon_start=1	/product="RNA helicase-DEAD box protein HHL16"
	/protein_id="AG54076.1"	/db_xref="GI:12621066"	/translation="MSGSTDENFRVLIISCPKRVKMLVEPLYDTLELPAPVKE OIORVATSGNMAVEELISTLEKGVHMLGTPREVENFHTGSLPARYKAPETLD PSPFENNADEYLOLNLLOPLTVKLIVRYLDKCMHITLTIDKRIIAVENKGN ESGVRELLKRIYQENKPSAPLNLRYGNELVQLSNCSESNALLENLSVDG OVEDLSTTVQPLEK EYVMENNSSESPDSVSVNITSLAEAGSLDESLGH NSNMGSDSGTSGSDSEVARYASPELOLRPOMEVLAIPALCKNLIICLPDGS KTRVAVYIAKDLHDKKKKASEKGVLYLVNKKVLEVDLKEPDLKKNVYGLSG DTOLKISPEYVYKSDIITSTOILENLIINLEMDPAVLSDSHLIDRCHTNK EAVYNNINRHYLMOKLKNRLEKKNRPILPDIIGLTNIGVGAFCOKAKAEHLK LCANDALFTIKVKNELDLKNOIQEPCKRAIADATRIPEKELILNIRIQYCO NSPMSDFQPYEOMAIOMEKKAKEGNKESVCAEHLKYNKALQINDIIMIDAVT HLETFYNEEKDKFAVIEDSDGDEDEDEDEDKPKLDEIDRELMITFE NNMKILRLAENPEYENELKLTNITIMQVTPTESSARLPTTRQVAVALSOMITE NEKFAVCYKAHLIGAGHSSFEKPTONFOKEVLSKSNVIREIVNDEHEKMTK KECNIVIRGLVTNEIYAVOAGARARADESTVLAHSSNVIETREIVNDEHEKMTK ALHCVONMKPEETARKILELOMOSIMEKIKTKRIAKYKNNPSLIPLCKNSVLA CSGEDIHVLEKMHVNMTPPEKELYIVRENKALQCKCAVYINGEILCKGAMGTMM YHKGIDLPCLKIRNFVVVFKNNSTKKYKKWELPTEINIDYSCILPDED"
misc_feature	1148..1201	/note="Region: DEAD-like helicase domain"	
misc_feature	1481..1627	/note="Region: DEAD/DEAH box"	
misc_feature	2606..2647	/note="conserved C-terminal helicase region: Region: helicase C"	
BASE COUNT	1158 a 637 c 748 g 830 t		
ORIGIN			
Query Match	99.2%; Score 3338.8; DB 9; Length 3373;		
Best Local Similarity	99.8%; Pred. No. 0;		
Matches 3343; Conservative	0; Mismatches 7; Indels 0; Gaps 0;		
15	GAGCCTGTGGACAACCTGCTATTTGTGAGGCAAGACAGCGGTGATGACCCCTTTCTTAAAG 74		
1	GGGCCCTGTGGACAACCTGCTATTTGTGAGGCAAGACAGCGGTGATGACCCCTTTCTTAAAG 60		
75	TGGGCAAGGAGACAGGCGGACGACATTTACCGTCTCGGACAGACAGCAATTCATCTGT 134		
61	TGGGCAAGGAGACAGGCGGACGACATTTACCGTCTCGGACAGACAGCAATTCATCTGT 120		
135	TGGGAGAACCTCTGCTCTCTGTGAGAAAGAAAGATGTGATGGTATTCATGACAGCA 194		
121	TGGGAGAACCTCTGCTCTCTGTGAGAAAGAAAGATGTGATGGTATTCATGACAGCA 180		
195	GAAATTCGGGTATCTATCTCGTGTGAGGCGGACAGGAGGAAATTTACCTCAGGTGCA 254		
181	GAAATTCGGGTATCTATCTCGTGTGAGGCGGACAGGAGGAAATTTACCTCAGGTGCA 240		
255	GGCTGTGTGAGTACCTGACCTTTCTGCTGACAGAGTGAAGAGACAGATTCAGAGAG 314		
241	GGCTGTGTGAGTACCTGACCTTTCTGCTGACAGAGTGAAGAGAGACATTCAGAGAG 300		
315	AGTGGCACCTCCGGGACATGACGACAGTGAATGCTGCTGAGACACATTCAGAGAGG 374		
301	AGTGGCACCTCCGGGACATGACGACAGTGAATGCTGCTGAGACACATTCAGAGAGG 360		
375	AGTGTGGACCTTGTGTGAGTGTGAGAAATTCGTGGAGGCGCTTCGGAGATTCGGCAGCC 434		
361	AGTGTGGACCTTGTGTGAGTGTGAGAAATTCGTGGAGGCGCTTCGGAGATTCGGCAGCC 420		
435	TCTGGGCGCGCTTACATGAACCTGAGCTCAAGGACCTTGCCTTCACATCTTTTGAGAA 494		
421	TCTGGGCGCGCTTACATGAACCTGAGCTCAAGGACCTTGCCTTCACATCTTTTGAGAA 480		
495	CGCTCATGATGATATCTCCAACTGCTGAACCTCTCTCAGGCACTCTTACAGAGCT 554		
481	CGCTCATGATGATATCTCCAACTGCTGAACCTCTCTCAGGCACTCTTACAGAGCT 540		
555	TCTAGTAGACGCTCTTGGATTAAGTCATGAGAGGAGAACTGTTGACATTTGAAATAG 614		
541	TCTAGTAGACGCTCTTGGATTAAGTCATGAGAGGAGAACTGTTGACATTTGAAATAG 600		
615	AAACCGGATTCGTCTGCGAGAAACCATGAAATTAATCAGGTGTAAGAACTTATTA 674		
601	AAACCGGATTCGTCTGCGAGAAACCATGAAATTAATCAGGTGTAAGAACTTATTA 660		
675	AAGGATTCGAGAAACAACTGCTCTGCAATTTCTGAAATGTTCTCGTAAACAGG 734		
661	AAGGATTCGAGAAACAACTGCTCTGCAATTTCTGAAATGTTCTCGTAAACAGG 720		
735	AAACATGACCTTGTCCAAAGATTAACGAGCTTATTCAGAAACCAATTCATAT 794		
721	AAACATGACCTTGTCCAAAGATTAACGAGCTTATTCAGAAACCAATTCATAT 780		
795	TGAGATTTATACAAAGTGTGTCGTCGCAAGTCGAGACGACACTTCTTCAGC 854		
781	TGAGATTTATACAAAGTGTGTCGTCGTCGCAAGTGTGAGACGACACTTCTTCAGC 840		
855	TCAGCCAAATCTGGAGAGGAGGTCTGGGGCATGGAGAAATACATCATGAAATCTTT 914		
841	TCAGCCAAATCTGGAGAGGAGGTCTGGGGCATGGAGAAATACATCATGAAATCTTT 900		
915	TGCATTTCTTGTGTAGTTTGAGAAATGACAGCAAGTTTGGCAGAGCAAGTGTACTG 974		
901	TGCATTTCTTGTGTAGTTTGAGAAATGACAGCAAGTTTGGCAGAGCAAGTGTACTG 960		
975	CTTAGATGAAGTCTTGGACATTAACGACACATGAGGAGTATTAAGCAATTCATAG 1034		
961	CTTAGATGAAGTCTTGGACATTAACGACACATGAGGAGTATTAAGCAATTCATAG 1020		
1035	TGATTCAGATGAAGAAATGTGGGACAGAAAGACATCCCGGACAGCAACTTCATAG 1094		
1021	TGATTCAGATGAAGAAATGTGGGACAGAAAGACATCCCGGACAGCAACTTCATAG 1080		
1095	GCCTTACCAAAATGGAAGTTGCCACACAGCTTGTGAAGGAAAGATATCATCAT 1154		
1081	GCCTTACCAAAATGGAAGTTGCCACACAGCTTGTGAAGGAAAGATATCATCAT 1140		
1155	CCCTACAGGAGTGGAAAAACACAGAGTGGCTGTATTCATTCGCAAGATCAATGACAA 1214		
1141	CCCTACAGGAGTGGAAAAACACAGAGTGGCTGTATTCATTCGCAAGATCAATGACAA 1200		
1215	GAAAGAAAAAGCATCTGAGCCTGGAAGAAATTAATAGTCTTGTCAATTAAGTAT 1274		
1201	GAAAGAAAAAGCATCTGAGCCTGGAAGAAATTAATAGTCTTGTCAATTAAGTAT 1260		
1275	TGAACAGCTCTCGGCAAGGAGTTCACACCAATTTTGAAGAAAGTGTGCTATTGG 1334		
1261	TGAACAGCTCTCGGCAAGGAGTTCACACCAATTTTGAAGAAAGTGTGCTATTGG 1320		
1335	ATTAAGTGTGATACCAACTGAAATATCATTTTCAGAAAGTGTGCAAGTCTGATAT 1394		
1321	ATTAAGTGTGATACCAACTGAAATATCATTTTCAGAAAGTGTGCAAGTCTGATAT 1380		
1395	TATTAATCAGTACGCTCAAAATCTTGAAGAACTCCTTTAACTTGAAGAAATGAGAA 1454		
1381	TATTAATCAGTACGCTCAAAATCTTGAAGAACTCCTTTAACTTGAAGAAATGAGAA 1440		
1455	TGCTGTGTGATGATGAGAAATTTTCCCTCATTTATCATGATGAGAAATGAGAA 1514		
1441	TGCTGTGTGATGATGAGAAATTTTCCCTCATTTATCATGATGAGAAATGAGAA 1500		
1515	CAAAAGACAGTGTATTAATTAACATCATGAGCATTAATTTGATGAGAAATGAGAA 1574		
1501	CAAAAGACAGTGTATTAATTAACATCATGAGCATTAATTTGATGAGAAATGAGAA 1560		
1575	TAGACTCAAGAAAGAAACCAAGATGATTCCTTCCTCAGATGAGGAGCTAAAGC 1634		
1561	TAGACTCAAGAAAGAAACCAAGATGATTCCTTCCTCAGATGAGGAGCTAAAGC 1620		
1635	TTCACCTGTGTGAGAGGCGCCAGACAGCAAGCCAAACCTGAGAGCAACATTTAAACT 1694		

OY	1797	TGCACCCAGGAAGAATCCATTTTTAAAGAAACTCTAAGAAAATAATGACAAAGATTTCAAAC	1865
Db	1866	TGACACCAGAGAAAAATCCATTTTAAAGAGAACTCTAGAAAATTTATGGCAAGCATTCACAC	1935
OY	1857	TTATTGTCAAATGAGTCCAAATGTCAGATTTTGGAACTCACAACCTTAATGAAATATGGCGCAT	1916
Db	1926	TTACTGCCAAAAAGTCCAAATGTCAATTTTGGAAACCCAACTATTATGAGTATGGCGCAT	1985
OY	1917	TCAAATGGAAAAAAAAGCTGCAAAAAAAGSAAATGGCAACAGACCTGTTTGTCAGAAC	1976
Db	1986	TCAAATGGAAAAAAAAGCTGCTTAAAGACGAAATGCCAAAGATCCAGTCTGTGTCAGAAC	2045
OY	1977	TTTAGAGAAAGTACAAATGAGGCCCTACAAATTTATGACAAATTCGAATGAGTATAGATGCTA	2036
Db	2046	TTTAGAGAAAGTACAAAGSAAATGCTCAATTTTGGAAACCCAACTATTATGAGTATGGCGCAT	2105
OY	2037	TACTCATCTTGAAACCTTTCTATATATGAGAGAAAGATATAGAAAGTTTGCATATAGAGA	2096
Db	2106	TAGCACCTTGAGAGCATTTCTACACTGATGAGAAAGAAAGAGTTCCGATCTCTCTCA	2162
OY	2097	TGATATGATGAGAGGCTGCTGATCATGATATGATATGATGCTGATGAAAGATATGAGATGAT	2156
Db	2163	TGACAGCGACAGAGGATATATACAGAGCCACAGCACTTTCATATGCCAACTTATAGGCGATCT	2222
OY	2157	AAAGAAACCTTGTGAACCTGSATGAAACAGATATGATTTTCATGACTTATTTTTGAAAA	2216
Db	2223	AAAGAAATCTTGTGAACCTGSAGCAAAACGSGATGATTTTCATGAAATTTGTTTGTATGA	2282
OY	2217	CAATTAATTTGTGAAAAAGCGTGGCTGAAAAACCCAGAAATATGAAATATGAAATGAAATGAA	2276
Db	2283	CAGAAAAATGTTGAAAAAACTACTCTAATAAAACCCAAATATGAGAAATGAAATATGAAATGAA	2342
OY	2277	ATTAGAGAAATACCATATATGAGAGCAATATACTAGAGATGAGAGATAGACAATAGAAATAT	2336
Db	2343	ATTAGAGAAACAGCATACTGSAACAAATTCACMAAGTGTGAGAGTCTCTCTCAGAGAAATAT	2402
OY	2337	CTTTACAAAAACACGACAGAGTGCATATGCGCTTCCAGTGAATTACTTAATATGAAAA	2366
Db	2403	TTTTACAAAAACACGACAGAGACACCTACGCACTTCCCAGTGAATATGTAATATGCAAA	2462
OY	2397	ATTGTGGAAGTATGAGATTCAAAGCCCAACCTCGATTTGAGTGGAGCAATATGAGTGAAT	2466
Db	2463	GTTTGGCGAAATTTGAGATTCAAAGCGATCACCCTGATTTGGCGGGGCAATATGAGTGAAT	2522
OY	2457	CAAAACCATGACACAGAAATGACAAAAAAGAAAGTCAATTAGTAATTTCCGTATGAAAAAT	2516
Db	2523	CAAGCCCATGATCAGATCAGATGACAAAAAAGAAAGTCAATTAGTAATTTCCGTATGCGAAT	2582
OY	2517	CAATCTGCTTATGCTTACCCACAGTGGCAGAGAAAGGTCTGGATATTAAATATATGTAACAT	2576
Db	2583	AAATCTGCTTATGCTTACCCACAGAGGTGGCAGAGAAAGGTCTGGATATTAAATATGTAACAT	2642
OY	2577	TGTTATCGGTATATGTTCTCTCCACCAATGAAATATAGCATGATGCTCAGAGGCTATGGTCGAGC	2636
Db	2643	TGTTATCGGTATATGCTTCTCTCCACCAAGAGATAGCCATGCTCAGAGGCTATGGTCGAGC	2702
OY	2637	CAGAGCTGATGAGACACCTTACGTCTGTTGCTACAGTGGTTCAGGAATATGAGAAC	2696
Db	2703	CAGAGCTGATGAGAAACAGTATGCTCTGCTCAGACACACAGTGGTTCAGGAATATGAGAAC	2762
OY	2697	TGACACAGTTATATATTTGAGAGAAAGATGATATTAAGGTATACATCTTGTTCCAAA	2756
Db	2763	GGACATGTTTATATATTTCCGAGAGAAAGATGATATTAAGGTATTAATCTTGTTCCAAA	2822
OY	2757	TATGAAACAGAGAGATGATCTCATATAGATTTTGGAAATTAACATGCAATATATATATGGA	2816
Db	2823	CATGAAACAGAGAGATGATCATATTAAGATTTTGGAAATTAACATGCAATATATATATGGA	2882
OY	2817	AAAGAAATGAAACCAAGAAATATTTGCCAAGCATTCAGAAATATATATATATATATATAT	2876
Db	2883	AAAGAAATGAAACCAAGAAAGCATTTCCAAGCAATTCAGAAATATATATATATATATATAT	2942

OY	2877	AACCTTCGTTGGCAAAAAGTGCAGTGCTCGACCGGTCTG3GGGAAGATATCATATGTAT	2936
Db	2943	AACACTTCTCIBCAAAAAATTGACATGCTGCTGCTGGGAGAAACATCTCATGTGAT	3002
OY	2937	TGAGAAAATGCATCACGTCAATATGACCAGAACATTCAGAACAATTACATTTAGAGA	2996
Db	3003	TGAGAAAGATGCATCATGTCCATATATGACACCAAGAAATCAAAGG9ACTGTACATTTAAGAGA	3062
OY	2997	AAACAAGACACTGCAAAAAGAAAGTGTGCCGACTATCAATAATAATGTTGAAATCATTTGCAA	3056
Db	3063	AAACAAGACACTGCAAAAAGAAATTTGCGATTATTCAGACCAATGAGAGATATATGCAA	3122
OY	3057	ATGTGGCCAGGCTGGGGGACAAATGATGATGTCGACAAAGGCTTAGATTGGCTTGCTGCA	3116
Db	3123	GTTGGCCAGGCTGGGGGACAAATGATGATGTCGACAAAGGCTTAGATTGGCTTGCTGCA	3182
OY	3117	AATAAGCAATTTTGTAGTGTTCGTTTTCAAAAATTAATTCACAAAGAAACAAATACAAAAAGTG	3176
Db	3183	AATAAGCAATTTTGTAGTGTCAATTTCAAAAATTAATTCACAGGAAAGAAAGCTACAAAGTG	3242
OY	3177	GGTACAAATTAACCTATACATTTCCCAATCTTGACTATTCAGATGCTGTTATTTAGTGA	3236
Db	3243	GGTGAATTTGCCATACAGATTTCTCGAATCTTGACACACAGATTCCTGTTAATCTGA	3302
OY	3237	TGAGATTAAGCACTGATTTGAAGATTTCTTTTAAATACATACAGTAAACATTTAAATAT	3296
Db	3303	TGAGATTTAGCACTGATTTGATGATTTATTTTAAATATCTGCTCCAACTCAACATTCATATG	3362
OY	3297	ATTATGATTTAATGATTAATTCATTATGCTACAGAACTGACATAGAAATCAATATAA	3348
Db	3363	ATTATTAATTTGTTGTTTCTGCTACACTGACACTTAAGCTTAAGATATATAGTA	3414
RESULT 5			
AY075132		3078 bp mRNA linear ROD 20-MAY-2002	
LOCUS		Mus musculus HELICARD mRNA, complete cds.	
DEFINITION		AY075132	
ACCESSION		AY075132.1 GI:18698980	
VERSION			
KEYWORDS		Mus musculus.	
SOURCE		Mus musculus.	
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
REFERENCE		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.	
AUTHORS		Kovacsosics,M., Martinon,F., Michéau,O., Bodmer,J.L., Hofmann,K.	
TITLE		overexpression of Helicard, a CARD-containing Helicase Cleaved	
JOURNAL		during Apoptosis; Accelerates DNA Degradation	
MEDLINE		Curr. Biol. 12 (10), 838-843 (2002)	
PUBMED		22009979	
REFERENCE		12015121	
AUTHORS		2 (bases 1 to 3078)	
TITLE		Kovacsosics,M., Hofmann,K. and Tschopp,J.	
JOURNAL		Direct Submission	
FEATURES		Submitted (24-JAN-2002) Biochemistry, University of Lausanne, ch	
source		des Boveresses 155, Epalinges, VD 1066, Switzerland	
		Location/Qualifiers	
		1..3078	
		/organism="Mus musculus"	
		/db_xref="taxon:10090"	
		1..3078	
		/note="helicase; contains two CARD domains"	
		/codon_start=1	
		/product="HELICARD"	
		/protein_id="AA177205.1"	
		/db_xref="GI:18698981"	
		/translation="MSIVCSAEDSFRLILFEPRLMYIIVEFLVLIHLIFLSAEKE	
		QILKIKNGCMTSAEELISTLBQGMPLGATOMFVALEHSGNPLARVKKPTLTD	
		ESPSERHADECHLLTLTIPTLVDKLINVDLTCEFGII IYEDFNRIISAAGNSGN	
		ESSGRFLRIRIVOKENMSTFDLVRONGNALREQLTGSGCPENDILANSRRDPP	
		AANCLLPAREDESLETETAMNVADLIDPEASDTDSVTTESTSIABSVCFOSDLGH	
		NSNMGRDGTMGSDSDSEVIDITKVSPPEPDLRPYQVEAQAPLDCKNIILCIPTGS	
CDS			


```

Db 1564 CACGAGAAATTCAGAGGAGCTCTACATGTAGAGAGAAACAAAGACAGTGCAGAGAAATTTG 1623
QY 3023 CCGACTATCAAAATTAATGTTGTAATATCATCTGCAGAAATGTCGAGGCTGTGCGAACATCA 3082
Db 1624 CTGATTTATGAGACATGAGAGATATATCTGCAGAGTGCAGAGCTGTGCGAACATCA 1683
QY 3083 TGGTGACAAAGGCTAGATTGCTGTCTCAAAATTAAGSAATTTGTGCGTTTCA 3142
Db 1684 TGGTGACAAAGGCTAGATTGCTGTCTCAAAATTAAGSAATTTGTGCGTTTCA 1743
QY 3143 AAAATATTCACAAAGAAACAAATACAAAAGTGGGTAGAAATTAACCTATCAATTTCGCA 3202
Db 1744 AAAATATTCACAGCAAGAAACAACTACAGAGAGTGGGTAGAAATTAACCTATCAATTTCG 1803
QY 3203 ATCTTACATCTACAGAAATGCTGTTTATTAGTATGAGATGAGATTAAGCACTCAATTGA 3262
Db 1804 ATCTTACATCTACAGAAATGCTGTTTATTAGTATGAGATGAGATTAAGCACTCAATTGA 1863
QY 3263 CTTTAAATTAATCTATCACTTAATCAATTAATGATTAAGTATTAATGATTAATGATTA 3322
Db 1864 ATTTTAAATTAATGCTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 1923
QY 3323 ACAGAACTGACATAGATCAATCAATAAA 3348
Db 1924 ACTGACCTAACTCTGATGATATGATGA 1949

RESULT 9
AX300838 1284 bp DNA line 1 PAT 30-NOV-2001
LOCUS Sequence 7 from Patent WO0185955.
DEFINITION AX300838
ACCESSION AX300838
VERSION AX300838.1 GI:17382116
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE
1 Bahr, G., Cocude, C. and Capron, A.
  Rh16 polypeptides and its fragments and polynucleotides encoding
  said polypeptides and therapeutic uses
  Patent: WO 0185955-A 7 15-NOV-2001;
  Istac (FR) : INSTITUT PASTEUR DE LILLE (FR)
FEATURES
  source 1..1284
    location/Qualifiers
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
BASE COUNT 489 a 198 c 261 g 335 t 1 oth is
ORIGIN
Query Match 37.3%; Score 1254.6; DB 6; Len 1284;
Best Local Similarity 99.6%; Pred. No. 1.8e-254;
Matches 1257; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 2103 TGAATGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2162
Db 1 TGAATGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 60
QY 2163 ACCTTGAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2222
Db 61 ACCTTGAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120
QY 2223 AATGTTGAAAAGGCTGCTGAAAACCCAGAAATATGAATGAAGAGCTGCTGAAAATTAAG 2282
Db 121 AATGTTGAAAAGGCTGCTGAAAACCCAGAAATATGAATGAAGAGCTGCTGAAAATTAAG 180
QY 2283 AATTAACATATGAGCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2342
Db 181 AATTAACATATGAGCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
QY 2343 AAAAAGCAGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2402

```

```

Db 241 AAAAAGCAGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
QY 2403 TGAATGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2462
Db 301 TGAATGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
QY 2463 CATGACAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2522
Db 361 CATGACAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
QY 2523 GCTTATGCTTACAGAGTGGCGAGAGAGAGTGTGATATTAAGAAATTAACATTAAT 2582
Db 421 GCTTATGCTTACAGAGTGGCGAGAGAGAGTGTGATATTAAGAAATTAACATTAAT 480
QY 2583 CCGTATGCTTACAGAGTGGCGAGAGAGAGTGTGATATTAAGAAATTAACATTAAT 2642
Db 481 CCGTATGCTTACAGAGTGGCGAGAGAGAGTGTGATATTAAGAAATTAACATTAAT 540
QY 2643 TGAATGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2702
Db 541 TGAATGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
QY 2703 AGTTAATGATTTCCGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2762
Db 601 AGTTAATGATTTCCGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
QY 2763 ACCAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2822
Db 661 ACCAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
QY 2823 AATGAAAACCAAGAGAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2882
Db 721 AATGAAAACCAAGAGAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
QY 2883 CCTTGGCAAAAACCTGAGTGTGCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2942
Db 781 CCTTGGCAAAAACCTGAGTGTGCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
QY 2943 AATGATCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3002
Db 841 AATGATCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
QY 3003 ACAGCTGCAAAAAGAGTGGCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3062
Db 901 ACAGCTGCAAAAAGAGTGGCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
QY 3063 CCAGGCTTGGGAGACAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3122
Db 961 CCAGGCTTGGGAGACAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1020
QY 3123 GAATTTTGTAGTGTGTTTAAAAATTAATTAACAAAGAAACAAATTAACAAATTAACAAATTAAC 3182
Db 1021 GAATTTTGTAGTGTGTTTAAAAATTAATTAACAAAGAAACAAATTAACAAATTAACAAATTAAC 1080
QY 3183 ATTACCTATCACTTTCCCAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3242
Db 1081 ATTACCTATCACTTTCCCAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
QY 3243 TTAGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3302
Db 1141 TTAGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
QY 3303 ATTAATGATTTCTTATGCTACAGAACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 3362
Db 1201 ATTAATGATTTCTTATGCTACAGAACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260
QY 3363 CT 3364
Db 1261 NT 1262

```

RESULT 10

BC007966
 LOCUS BC007966 1213 bp mRNA
 DEFINITION Homo sapiens, clone IMAGE:4304805, mRNA.
 ACCESSION BC007966
 VERSION BC007966.1 GI:14044090
 KEYWORDS
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 1213)
 REFERENCE
 TITLE Direct Submission
 AUTHORS Strausberg, R.
 JOURNAL Submitted (11-MAY-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 COMMENT
 Contact: MGC help desk
 Email: cgapbe-rt@mail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: <http://www.nisc.nih.gov/>
 Contact: nisc.mgc@nrl.nih.gov
 Sherchenko, Y., Metherby, K.D., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Briley, C., Brooks, S., Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Kalns, E., Legaspi, R., Lim, M., Maduro, Q.L., Mastello, C., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Snyder, B., Stantiford, S., Thomas, P.J., Tjongson, E.E., Touchman, J.W., Tsurganov, C., Vogt, L.L., Walker, M.A., Zhang, L.-R. and Green, E.D.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IMAGE:4304805
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.
 FEATURES
 source location/Qualifiers
 1..1213
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4304805"
 /issue_type="Uterus, leiomyosarcoma"
 /clone_id="NIH_MGC_46"
 /lab_host="DH10B-R"
 /note="Vector: pORF7"
 BASE COUNT 386 a 247 c 285 g 295 t
 ORIGIN
 Query Match 35.5% Score 1194.4; DB 9; Length 1213;
 Best Local Similarity 99.5%; Pred. No. 9,66-242;
 Matches 1198; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

620 GGATTGCTGTCAGAAAAAATGAAATGATCAGTGTAAAGAGCTACTAAAGAGA 679
 249 GGATTGCTGTCAGAAAAAATGAAATGATCAGTGTAAAGAGCTACTAAAGAGA 308
 660 TTGTGCAGAAAGAAAGCTGTTCTGTCATTTGCAATGTTCTTGTCAACAG:AAACA 739
 309 TTGTGCAGAAAGAAAGCTGTTCTGTCATTTGCAATGTTCTTGTCAACAG:AAACA 368
 740 ATGAACTGTCCAAAGTAAGCTGCTGATTTGCTGGAAGAAAGCAATGATTAAGA 799
 369 ATGAACTGTCCAAAGTAAGCTGCTGATTTGCTGGAAGAAAGCAATGATTAAGA 428
 800 ATTATACCAAGTTGATGTTGCTGCTCAAGTGAAGAGCAATTTTTCACACAGAT:AGC 859
 429 ATTATACCAAGTTGATGTTGCTGCTCAAGTGAAGAGCAATTTTTCACACAGAT:AGC 488
 860 CAATCTGGAGAGAGAGCTCTGGGCGATGAGAGAAATACCTCATCATCTTTTCAG 919
 489 CAATCTGGAGAGAGAGCTCTGGGCGATGAGAGAAATACCTCATCATCTTTTCAG 548
 920 ATTCTTGTAGTTTTCAGATTCAGACAGAGTTTGGCAGAGAGAGTGTAGCTG:TTAG 979
 549 ATTCTTGTAGTTTTCAGATTCAGACAGAGTTTGGCAGAGAGAGTGTAGCTG:TTAG 608
 980 ATGAAAGTCTTGGACATTAACAGCAACATGGGCACTGATTCAGGCACCATGGGAAGTAT 1039
 609 ATGAAAGTCTTGGACATTAACAGCAACATGGGCACTGATTCAGGCACCATGGGAAGTAT 668
 1040 CAGATGAGAGAAATGTGGCAGCAGAGATGATCCCGGAGCCAGAAATTCAGTCAAGGCTT 1099
 669 CAGATGAGAGAAATGTGGCAGCAGAGATGATCCCGGAGCCAGAAATTCAGTCAAGGCTT 728
 1100 ACCAATGGAAGTTTCCCGAGCAGCTTGAAGAGGAAGAAATATCATCTTGCCTCCTTA 1159
 729 ACCAATGGAAGTTTCCCGAGCAGCTTGAAGAGGAAGAAATATCATCTTGCCTCCTTA 788
 1160 CAGGAGTGGAAAAACAGAGTGGCTTTTACATTTGCCAGAGTACTTACACAGAGAA 1219
 789 CAGGAGTGGAAAAACAGAGTGGCTTTTACATTTGCCAGAGTACTTACACAGAGAA 848
 1220 AAAAAAGCATGAGCTGAGGAAAAAGTTATAGTTCTTGAATATAGTCTGTATAGTAA 1279
 849 AAAAAAGCATGAGCTGAGGAAAAAGTTATAGTTCTTGAATATAGTCTGTATAGTAA 908
 1280 AGCTTCTGGCAGAGAGTTCACACATTTTGAAGAAATGTATGCTGTATATTA 1339
 909 AGCTTCTGGCAGAGAGTTCACACATTTTGAAGAAATGTATGCTGTATATTA 968
 1340 GTGTGATACCCCACTGAATATCATTTCCAGAAATTTGCAAGTCTGTATATTA 1399
 969 GTGTGATACCCCACTGAATATCATTTCCAGAAATTTGCAAGTCTGTATATTA 1028
 1400 TCAGTACAGCTCAAAATCCCTTGAAGAAATCCCTTGAAGAAATGTATATTA 1459
 1029 TCAGTACAGCTCAAAATCCCTTGAAGAAATCCCTTGAAGAAATGTATATTA 1088
 1460 GTGTTCATTTGTCAGACTTTTCCCTCATATTCATTTGAGATGATGATCATCA:CAAA 1519
 1089 GTGTTCATTTGTCAGACTTTTCCCTCATATTCATTTGAGATGATGATCATCA:CAAA 1148
 1520 AAGCAGTATATTAATCATCATGAGGATTAATTTGATGACAGAGTGAAGAAAT:AAAG 1579
 1149 AAGCAGTATATTAATCATCATGAGGATTAATTTGATGACAGAGTGAAGAAAT:AAAG 1208
 1580 TCMA 1583
 1209 AAAA 1212
 RESULT 11
 BC004031 2050 bp mRNA linear 07-AUG-2002
 LOCUS BC004031

Query Match	23.9%	Score 803;	DB 6;	length 1258;
Best Local Similarity	95.5%	Pred. No. 4.5e-15;		
Matches 853; Conservative	0;	Mismatches 0;	Indels 40;	Gaps 1;

[illegible]

AUTHORS Peyman, J.A., da Silva, A., Hochman, P. and Hsu, A.
 TITLE Interferon induced polynucleotides and proteins encoded thereby
 Patent: WO 0118208-A 11-15-MAR-2001
 JOURNAL Curagen Corporation (US); BIOGEN, INC. (US)

FEATURES
 source Location/Qualifiers

1..1270
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1..927
 /note="unnamed protein product"

CDS

/codon_start=1
 /protein_id="CAC35442.1"
 /db_xref="GI:13515372"
 /translation="MRAAMVNSNTPRRDVLSKPPVAG:RPPCESKRAADSGTHISPV
 POTAPSMENPLPSLRKDYEMVFRREPLSHLVLDY:YSENVHPRKEQIQRTVAT
 SGNOAVELLSTLEKGYWHGLQREFEALRRTGSP:AGRNKVELLDPSPEENA
 HDEYLDLNLPTLVLDKLVLDYLDKIMEELLIIEDNRNIAAENNGNSGYELL
 KRIYKEMFSAFLVRLROTGNNELVCLTSDCSSESNAH:ENLSQYDGPVEEQILS
 TVVQNLKEWGMENNSSESSFADSSVVSGL"

BASE COUNT 401 a 271 c 279 g 319 t

ORIGIN

Query Match 23.9%; Score 803; DB 6; length 1270;

Best Local Similarity 95.5%; Pred. No. 4,5e-159;

Matches 853; Conservative 0; Mismatches 0; Indels 40; Gaps 1;

45 GCACAGAGGGGTAGACCCCTCTCTAGTGGCGAGCGGAA:ACGGCAATTTCA 104
 67 GCACAGAGGGGTAGACCCCTCTCTAGTGGCGAGCGGAA:ACGGCAATTTCA 126
 105 CCTGTCCCGCAGACACACACCATCTGCTGGAGAACCTCTGCTTCTGAGAAG 164
 127 CCTGTCCCGCAGACACACACCATCTGCTGGAGAACCTCTGCTTCTGAGAAG 186
 165 AAGATGTGGAATGGTATTCCACAGAGAGAAATTTCCCTATTCATCTGCTTAC 224
 187 AAGATGTGGAATGGTATTCCACAGAGAGAAATTTCCCTATTCATCTGCTTAC 246
 225 GGCGAGGGTGAATGTACATCCAGGTGGAGGCTGTGCTGGAG:ACCTGAGTTTTC 284
 247 GGCGAGGGTGAATGTACATC----- 268
 285 TGCAGAGGTGAAGAGACAGATTTCAGAGAGAGTGGCCACTCTGAGAA:ACAGGCACT 344
 269 --CAGAGGTGAAGAGACAGATTTCAGAGAGAGTGGCCACTCTGAGAA:ACAGGCACT 326
 345 TGAAGTGTGCTGAGACCTTGGAGAGAGGAGTCTGGCA:CTTGTGATTCGGCAAT 404
 327 TGAAGTGTGCTGAGACCTTGGAGAGAGGAGTCTGGCA:CTTGTGATTCGGCAAT 386
 405 CGTGGAGGGCTCCGGAGAAACCGGCGCTCTGGCGCG:CCCGTACATGAA:CCCTGAGCT 464
 387 CGTGGAGGGCTCCGGAGAAACCGGCGCTCTGGCGCG:CCCGTACATGAA:CCCTGAGCT 446
 465 CAGGACTTGGCCCTTCATCTGTTAGAGACGCTCATGAAATATCTG:AACTGCTGAA 524
 447 CAGGACTTGGCCCTTCATCTGTTAGAGACGCTCATGAAATATCTG:AACTGCTGAA 506
 525 CCTCTCTCAGCCCTCTGCTGCTGAGCAAGCTTCTAGTTAGAGAT:CTTGTGATTCGGCAAT 584
 507 CCTCTCTCAGCCCTCTGCTGCTGAGCAAGCTTCTAGTTAGAGAT:CTTGTGATTCGGCAAT 566
 585 GGAGAGGAACTGTGACATTTGAAGAGAAACCGGATTCGTGTGCA:AAACAAATG 644
 567 GGAGAGGAACTGTGACATTTGAAGAGAAACCGGATTCGTGTGCA:AAACAAATG 626
 645 AATGATAGAGTGTAGAGAGCTACTAAAGATTTGAGAGAGAA:AAATGCTTCTG 704
 627 AATGATAGAGTGTAGAGAGCTACTAAAGATTTGAGAGAGAA:AAATGCTTCTG 686
 705 TGAATTTGGAATGTTCTTCTGTTAAACAGAGAAATGAACTT:CTTCAAA:CTTAA:AGG 764
 687 TGAATTTGGAATGTTCTTCTGTTAAACAGAGAAATGAACTT:CTTCAAA:CTTAA:AGG 746

765 CTCTGATTTGCTAGAAAGCAATGACAGATTGAGAAATTTATCACAAGTTGATGCTCA 824

747 CTCTGATTTGCTAGAAAGCAATGACAGATTGAGAAATTTATCACAAGTTGATGCTCA 806

825 AGTGAAGAGCAACTTCTTTCACACAGATTGAGCAAAATGTGAGAGAGAGTGTGGG 884

807 AGTGAAGAGCAACTTCTTTCACACAGATTGAGCAAAATGTGAGAGAGAGTGTGGG 866

885 CATGAGAGTAATCATCATGAGATTCATCTTTTGGAGATTTCTGTAGTTTCA 937

867 CATGAGAGTAATCATCATGAGATTCATCTTTTGGAGATTTCTGTAGTTTCA 919

RESULT 14

AX098234 3692 bp DNA linear PAT 30-MAR-2001

LOCUS Sequence 9 from Patent WO0118208.

DEFINITION AX098234

ACCESSION AX098234.1 GI:13515368

VERSION

KEYWORDS

SOURCE

ORGANISM

human.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 3692)

Peyman, J.A., da Silva, A., Hochman, P. and Hsu, A.
 Interferon induced polynucleotides and proteins encoded thereby
 Patent: WO 0118208-A 9-15-MAR-2001;
 Curagen Corporation (US); BIOGEN, INC. (US)

LOCATION/Qualifiers

1..3692

/organism="Homo sapiens"

/db_xref="taxon:9606"

1..804

/note="unnamed protein product"

/codon_start=1

/protein_id="CAC35442.1"

/db_xref="GI:13515369"

/translation="MRAAMVNSNTPRRDVLSKPPVAG:RPPCESKRAADSGTHISPV
 POTAPSMENPLPSLRKDYEMVFRREPLSHLVLDY:YSENVHPRKEQIQRTVAT
 SGNOAVELLSTLEKGYWHGLQREFEALRRTGSP:AGRNKVELLDPSPEENA
 HDEYLDLNLPTLVLDKLVLDYLDKIMEELLIIEDNRNIAAENNGNSGYELL
 KRIYKEMFSAFLVRLROTGNNELVCLTSDCSSESNAH:ENLSQYDGPVEEQILS
 TVVQNLKEWGMENNSSESSFADSSVVSGL"

BASE COUNT 1166 a 709 c 806 g 1011 t

ORIGIN

Query Match 19.7%; Score 663; DB 6; length 3692;

Best Local Similarity 92.9%; Pred. No. 1.7e-129;

Matches 722; Conservative 0; Mismatches 15; Indels 40; Gaps 1;

45 GCACAGAGGGGTAGACCCCTCTCTAGTGGCGAGCGGAA:ACGGCAATTTCA 104
 55 GCACAGAGGGGTAGACCCCTCTCTAGTGGCGAGCGGAA:ACGGCAATTTCA 114
 105 CCTGTCCCGCAGACACACACCATCTGCTGGAGAACCTCTGCTTCTGAGAAG 164
 115 CCTGTCCCGCAGACACACACCATCTGCTGGAGAACCTCTGCTTCTGAGAAG 174
 165 AAGATGTGGAATGGTATTCCACAGAGAGAAATTTCCCTATTCATCTGCTTCA 224
 175 AAGATGTGGAATGGTATTCCACAGAGAGAAATTTCCCTATTCATCTGCTTCA 234
 225 GGCGAGGGTGAATGTACATCCAGGTGGAGGCTGTGCTGAGACTACCTTGTGCG 284
 235 GGCGAGGGTGAATGTACATC----- 256
 285 TGCAGAGGTGAAGAGAGATTTCAGAGAGAGTGGCCACTCTGGAGACATGCAAG 344
 257 --CAGAGGTGAAGAGAGATTTCAGAGAGAGTGGCCACTCTGGAGACATGCAAG 314
 345 TGAAGTGTGCTGAGACCTTGGAGAGAGGAGTCTGGCACTTGTGATTCGGCAAT 404

Db 315 TGAACCTGCTGAGACCTTGGACAGAGGAGCTTGGACCTTTGTTGATTCGGGAAIT 374
 Oy 405 CGTGAAGGCTCCCGAGAGAACCGGAGCCCTCTGGCCCGCCGCTACATATTA 464
 Db 375 CGTGAAGGCTCCCGAGAGAACCGGAGCCCTCTGGCCCGCCGCTACATATTA 434
 Oy 465 CAGGAGCTGGCTCTGCAATCTTTGAGAAAGCTCATATGAAATCTG 524
 Db 435 CAGGAGCTGGCTCTGCAATCTTTGAGAAAGCTCATATGAAATCTG 494
 Oy 525 CCTCTTCAAGCCCTCTGCTGAGACAGCTCTAGTACAGAGCTTGA 584
 Db 495 CGCTCTTCAAGCCCTCTGCTGAGACAGCTCTAGTACAGAGCTTGA 554
 Oy 585 GAGAGGAGAACTGTTGACAAATTTGAAAGAGAAACCGGATTTGCTGCA 644
 Db 555 GAGAGGAGAACTGTTGACAAATTTGAAAGAGAAACCGGATTTGCTGCA 614
 Oy 645 AATGAATCAGGTGTGAAGAGAGCTACTAAAGAGATTTGCAAGAAAGATGCTTC 704
 Db 615 AATGAATCAGGTGTGAAGAGAGCTACTAAAGAGATTTGCAAGAAAGATGCTTC 674
 Oy 705 TGCATTTCTGAATGTTCTTCTGTCACAGAGAAACATGCTTCCCA 764
 Db 675 TGCATTTCTGAATGTTCTTCTGTCACAGAGAAACATGCTTCCCA 734
 Oy 765 CTCTGATTTGCTCAGAAAGAGATGAGATTTGCAATGATGATGCTTC 821
 Db 735 CTCTGATTTGCTCAGAAAGAGATGAGATTTGCAATGATGATGCTTC 791

RESULT 15
 AX098238 3704 bp DNA 11mb PAT 30-MAR-2001
 LOCUS AX098238
 DEFINITION Sequence 13 from Patent WO0118208.
 ACCESSION AX098238
 VERSION AX098238.1 GI:13515375
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 3704)
 AUTHORS Peyman, J.A., da Silva, A., Hochman, P. and Hsu, A.
 TITLE Interferon induced polynucleotides and proteins encoded thereby
 JOURNAL Patent: WO 0118208-A 13 15-MAR-2001;
 Curegen Corporation (US); BIOGEN, INC. (US)
 FEATURES
 source location/Qualifiers
 1..3704
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1..816
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC35443.1"
 /db_xref="GI:13515375"

CDS
 1..816
 /translation="MPAAVMSNTTTRDVIKSPVAV-GRPCFSKWAADSGTHISPV
 PQTAPSAMENPLPSLRKKDVEWFERREPLSHLYD-AGEVHPEVKEQIQRTVAT
 SGNQAVELLSTLEKGYHGLGTRFEVALRRTSP-LAAKMNPELDELPSFENA
 HDEYLLNLLOPTLVKDLIVRDVLDKMEFEILLITEDPHRILAAFNNGNSGYRELL
 KRIYKEMFSAFLNVLROTNNEIYVELTSDCSNACIGNTERDSNSA"
 BASE COUNT 1168 a 713 c 809 g 1014 t
 ORIGIN

Query Match 19.7% Score 663; DB 6; Length 3704;
 Best Local Similarity 92.9% Pred No. 1.7e-129;
 Matches 722; Conservative 0; Mismatches 15; Indels 40; Gaps 1;

Oy 45 GCACAGAGCGGTAGACCTGCTTCTTAAGTGGAGCGACAGCGGCAATTCATTC 104
 Db 67 GCACAGAGCGGTAGACCTGCTTCTTAAGTGGAGCGAGCGGCAATTCATTC 126
 Oy 105 CCGTCCCGCAGACAGACACCATCTGCTGGGAGAGCTCTCTCCTTTGTGAGAAAG 164

Db 127 CCTGTCCGCGAGACACAGACACCATCTGCTGGAGAAACCTCTCCCTCTCTAGAAAG 186
 Oy 165 AAAGATGTGGAATGGATTTTCACAGACAGAGAAATTTGCGATATCATCTGCTTCAG 224
 Db 187 AAAGATGTGGAATGGATTTTCACAGACAGAGAAATTTGCGATATCATCTGCTTCAG 246
 Oy 225 GCGCAGGATGAAAATGTATCATCCAGAGTGGAGCCCTGTGCTGATCACTGACCTTTCTGCC 284
 Db 247 GCGCAGGATGAAAATGTATCATC----- 268
 Oy 285 TGCAGAGTGAAGAGACGATTCAGAGACAGTGGCCACCTCCGGGAACATGACAGAGT 344
 Db 269 --CAGAGTGAAGAGACGATTCAGAGACAGTGGCCACCTCCGGGAACATGACAGAGT 326
 Oy 345 TGAACCTGCTGAGACACCTTGGAGAGAGAGTCTGGACCTTGTGCTGAGCTCGGAAAT 404
 Db 327 TGAACCTGCTGAGACACCTTGGAGAGAGAGTCTGGACCTTGTGCTGAGCTCGGAAAT 386
 Oy 405 CGTGAAGGCTCCCGAGAGAACCGGACGCTCTGAGCGCCGCTACATGAACCT 464
 Db 387 CGTGAAGGCTCCCGAGAGAACCGGACGCTCTGAGCGCCGCTACATGAACCT 446
 Oy 465 CAGGAGCTGGCTCTGCTGATGAGAGAGCTCATGATGATATGCTCACT 524
 Db 447 CAGGAGCTGGCTCTGCTGATGAGAGAGCTCATGATGATATGCTCACT 506
 Oy 525 CCTCTTCAAGCCCTCTGCTGAGACAGCTTCTAGTGAAGAGCTTGA 584
 Db 507 CCTCTTCAAGCCCTCTGCTGAGACAGCTTCTAGTGAAGAGCTTGA 566
 Oy 585 GAGAGGAGAACTGTTGACAAATTTGAAAGAGAAACCGGATTTGCTGCA 644
 Db 567 GAGAGGAGAACTGTTGACAAATTTGAAAGAGAAACCGGATTTGCTGCA 626
 Oy 645 AATGAATCAGGTGTGAAGAGAGCTACTAAAGAGATTTGCAAGAAAGATGCTTC 704
 Db 627 AATGAATCAGGTGTGAAGAGAGCTACTAAAGAGATTTGCAAGAAAGATGCTTC 686
 Oy 705 TGCATTTCTGAATGTTCTTCTGTCACAGAGAAACATGCTTCCCAAGATTAACAGG 764
 Db 687 TGCATTTCTGAATGTTCTTCTGTCACAGAGAAACATGCTTCCCAAGATTAACAGG 746
 Oy 765 CTCTGATTTGCTCAGAAAGAGATGAGATTTGCAATGATGATGCTTC 821
 Db 747 CTCTGATTTGCTCAGAAAGAGATGAGATTTGCAATGATGATGCTTC 803

Search completed: May 16, 2003, 18:23:30
 Job time : 8695 secs


```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5786526"
/clone_id="NH_MGC_71"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pCMV-Sp16; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.1 kb."
BASE COUNT      383 a      161 c      229 g      240 t
ORIGIN

```

```

Query Match      21.2% Score 714; DB 14; Length 1013;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 764; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

OY 1752 GAAAAACCAATACAGAGCCATGCAGAAAGTTGCCATTGCAGATGCAGACAGAGA 1811
DB 1 GAAAAACCAATACAGAGCCATGCAGAAAGTTGCCATTGCAGATGCAGACAGAGA 60
OY 1812 TCCATTAAAGAGAACTCTAGAAATATGACAAAGATTCAACTATTCTCAATGAG 1871
DB 61 TCCATTAAAGAGAACTCTAGAAATATGACAAAGATTCAACTATTCTCAATGAG 120
OY 1872 TCCAAATGCAGATTTTGGAACTCAACCTATGAACAAATGGCCATTCAATGSA 1931
DB 121 TCCAAATGCAGATTTTGGAACTCAACCTATGAACAAATGGCCATTCAATGSA 180
OY 1932 AGCTGCAAAAAAGAAATCGCCAAAGAAAGCTGTTGTCAGAACATTTGAGAAATGCA 1991
DB 181 AGCTGCAAAAAAGAAATCGCCAAAGAAAGCTGTTGTCAGAACATTTGAGAAATGCA 240
OY 1992 TGAGGCCCCATCAATATATGACAACTTCGATATGATATGATATGATATGAT 2051
DB 241 TGAGGCCCCATCAATATATGACAACTTCGATATGATATGATATGATATGAT 300
OY 2052 TTTCTATATGAGAGAAAGATTAAGAAAGTTGCAATGATGATGATGATGATG 2111
DB 301 TTTCTATATGAGAGAAAGATTAAGAAAGTTGCAATGATGATGATGATGATG 360
OY 2112 TGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2171
DB 361 TGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420
OY 2172 ACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2231
DB 421 ACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
OY 2232 AAGGCTGCTGAGAAACCCAGAAATGAAAGAGTGAATTAATTAACAAATACCAT 2291
DB 481 AAGGCTGCTGAGAAACCCAGAAATGAAAGAGTGAATTAATTAACAAATACCAT 540
OY 2292 AATGAGCAATATAGTACGAGTACGAGATGAGAGCAATTAATCTTTTAAACACG 2351
DB 541 AATGAGCAATATAGTACGAGTACGAGATGAGAGCAATTAATCTTTTAAACACG 600
OY 2352 ACAGAGTCATATGAGGCTTCCGAGTGTAGTGAATTAATTAATTTTGTAGTAGG 2411
DB 601 ACAGAGTCATATGAGGCTTCCGAGTGTAGTGAATTAATTAATTTTGTAGTAGG 660
OY 2412 AGTCAAAGCCAGCATCTGATTTGGAGTGCAGACAGAGTGAATGCAATGATGACACA 2471
DB 661 AGTCAAAGCCAGCATCTGATTTGGAGTGCAGACAGAGTGAATGCAATGATGACACA 720
OY 2472 GAATGACAAAAAGAGTATTAGTAAATTTGCGACTGCGAAAT 2516
DB 721 GAATGACAAAAAGAGTATTAGTAAATTTGCGACTGCGAAAT 765

```

```

RESULT 2
B0772836/c B0772836 755 bp mRNA 11base EST 26-JUL-2002
DEFINITION UI-H-FEO-bdn-c-04-0-UI.s1 NCI_CGAP_FEO Human sapiens cdna clone

```

ACCESSION VERSION KEYWORDS SOURCE ORGANISM

UI-H-FEO-bdn-c-04-0-UI 3', mRNA sequence.
B0772836
B0772836.1 GI:21981312
EST.
human.
Homo sapiens

REFERENCE AUTHORS TITLE

1 (bases 1 to 755)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL COMMENT

Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: James Martin
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, Bento-Soares@uiowa.edu
Seq primer: M13 FORWARD
POLY(A)-Yes.

FEATURES source

Location/Qualifiers
1..755
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-H-FEO-bdn-c-04-0-UI"
/clone_lib="NCI_CGAP_FEO"
/tissue_type="Chondrosarcoma Cell line"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoR I; Site_2: Not I; NCI_CGAP_FEO is
a CDNA library containing the following tissue(s): a pool
of 3 chondrosarcoma cell lines (grade 2) The library was
constructed according to Bonaldi, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT73-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(d)18 tail. The sequence tag for this library is
CGCTACGGAC. The cell lines was provided by Dr James Martin
of University of Iowa.
TAG_LIB=UI-H-FEO
TAG_TISSUE=Human grade 2 chondrosarcoma cell line pool
TAG_SEQ=CGCTACGGAC"
BASE COUNT 207 a 141 c 115 g 290 t 2 others
ORIGIN

```

Query Match      20.1%; Score 676; DB 14; Length 745;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 676; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 2690 TCGACATGAGACAGTAAATGATTTCCGAGAGAAAGATGATTAAGCAATATTTG 2749
DB 695 TCGACATGAGACAGTAAATGATTTCCGAGAGAAAGATGATTAAGCAATATTTG 636
OY 2750 TTCAAAATATGAACCGAGAGATGCTCATTAAGATTTTGGATATACAGATCAAAAGTA 2809
DB 635 TTCAAAATATGAACCGAGAGATGCTCATTAAGATTTTGGATATACAGATCAAAAGTA 576
OY 2810 TAAATGAAAAAGAAATGAACCAAGAAATATTCGACAGATTCACAGAAATACCAT 2869
DB 575 TAAATGAAAAAGAAATGAACCAAGAAATATTCGACAGATTCACAGAAATACCAT 516
OY 2870 CACTAATTAACCTCTTCTGCAAAAACAGTGTGCTAGCGCTTTTGGCGCAATATATTC 2929
DB 515 CACTAATTAACCTCTTCTGCAAAAACAGTGTGCTAGCGCTTTTGGCGCAATATATTC 456

```

QY 2930 ATGTATTTAGAAAAATGATCATGATATATGACCCAGAAATTTAGAAATTTACATTG 2989
 DB 455 ATGTATTTAGAAAAATGATCATGATATATGACCCAGAAATTTAGAAATTTACATTG 336
 QY 2990 TAAGAGAAACAAAGCACTGCAAAAAGAGTGTGCCACTATCAATTAATGCTGAAATCA 3049
 DB 395 TAAGAGAAACAAAGCACTGCAAAAAGAGTGTGCCACTATCAATTAATGCTGAAATCA 336
 QY 3050 TCTGCAATTTGCGCCAGCTGTGGGGACAAATGATGTGTCACAAAGGCTTACATTTGCTT 3109
 DB 335 TCTGCAATTTGCGCCAGCTGTGGGGACAAATGATGTGTCACAAAGGCTTACATTTGCTT 276
 QY 3110 GTCCTAAATTAAGAAATTTTGTAGTGTCTTCCAAAATTAATTCACAAAGCAATACCA 3169
 DB 275 GTCCTAAATTAAGAAATTTTGTAGTGTCTTCCAAAATTAATTCACAAAGCAATACCA 216
 QY 3170 AAAAGTGGTAGAATTTACCTATACATTTCCCAATCTTGACATTCAGAAATGCTGTTAT 3229
 DB 215 AAAAGTGGTAGAATTTACCTATACATTTCCCAATCTTGACATTCAGAAATGCTGTTAT 156
 QY 3230 TTAGATGAGAGATTAAGCACTGATGAAATCTTTTAAATTAATTCATCATTTAAACAT 3289
 DB 155 TTAGATGAGAGATTAAGCACTGATGAAATCTTTTAAATTAATTCATCATTTAAACAT 96
 QY 3290 TAAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGAT 3349
 DB 95 TAAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGAT 36
 QY 3350 TGATTTGTTTACTCTG 3365
 DB 35 TGATTTGTTTACTCTG 20

RESULT 3
 BM467983 1115 bp mRNA lineal EST 05-FEB-2002
 LOCUS AGENCOURT_6437921 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5532884
 DEFINITION 5', mRNA sequence.
 ACCESSION BM467983
 VERSION BM467983.1 GI:18517025
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 1115)
 NIH-MGC http://mgi.ncl.nih.gov/.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cstraub@mail.nih.gov
 Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: LLM12216 row: n column: 21
 High quality sequence stop: 690.

FEATURES

Location/Qualifiers
 1..1115

BASE COUNT 326 a 221 c 218 g 277 t 73 others
 ORIGIN
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5532884"
 /clone_id="NIH_MGC_71"
 /tissue_type="leiomysarcoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: uterus; Vector: pCMV-SPO6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: oligo dt.
 Average insert size 2.1 kb.

Query Match 19 68; Score 658; DB 13; Length 1115;
 Best Local Similarity 99.68; Pred. No. 0;
 Matches 808; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2555 TGGATTTTAAAGATGTATACATTTGTTATCCGTTATGCTGTGTCACCAATGAATATTA 2614
 DB 2 TGGATTTTAAAGATGTATACATTTGTTATCCGTTATGCTGTGTCACCAATGAATATTA 61
 QY 2615 TGTGTCAGGCCCCGTGTGTCAGCCAGACCTGATGAGACCTAGCTCTG;TATGACA 2674
 DB 62 TGTGTCAGGCCCCGTGTGTCAGCCAGACCTGATGAGACCTAGCTCTG;TATGACA 121
 QY 2675 GTGGTTCAGAGTATGCAACATGACAGACGTATATGATTTCCGAGAGAAATGAGATATA 2734
 DB 122 GTGGTTCAGAGTATGCAACATGACAGACGTATATGATTTCCGAGAGAAATGAGATATA 181
 QY 2735 AAGCTATACATTTGTTCAAAATATGAACAGAGAGATATGCTCATAGATTTGGAAT 2794
 DB 182 AAGCTATACATTTGTTCAAAATATGAACAGAGAGATATGCTCATAGATTTGGAAT 241
 QY 2795 TACAGATGCAAGATATATGAGAAAAGAAATGAACCAAGAGAAATATTGCTTAACAT 2854
 DB 242 TACAGATGCAAGATATATGAGAAAAGAAATGAACCAAGAGAAATATTGCTTAACAT 301
 QY 2855 ACAGAAATTAACCCATCACTATTAATCTTCTGCAAAAACAGAGTGTGTAACCTGTT 2914
 DB 302 ACAGAAATTAACCCATCACTATTAATCTTCTGCAAAAACAGAGTGTGTAACCTGTT 361
 QY 2915 CTGGGGAAGATATCCATGTATATGAGAAAATGCATCACTGATATATGACCCATATTA 2974
 DB 362 CTGGGGAAGATATCCATGTATATGAGAAAATGCATCACTGATATATGACCCATATTA 421
 QY 2975 AGCACTTTACATTTGTAAGAGAAAACAAAGCACTGCAAAAAGATGTGCCGATATATAA 3034
 DB 422 AGCACTTTACATTTGTAAGAGAAAACAAAGCACTGCAAAAAGATGTGCCGATATATAA 481
 QY 3035 TAAATGGTGAATCATGTCGCAATATGTGGCCAGCTTGAGGAAACAATGATGTTGACAAAG 3094
 DB 482 TAAATGGTGAATCATGTCGCAATATGTGGCCAGCTTGAGGAAACAATGATGTTGACAAAG 541
 QY 3095 GCTTAGATTTGCTGTCGTCGCAAAATTAAGAAATTTGTAGTGTGTTTCAAAAATTAATTA 3154
 DB 542 GCTTAGATTTGCTGTCGTCGCAAAATTAAGAAATTTGTAGTGTGTTTCAAAAATTAATTA 601
 QY 3155 CAAAGAAACATTAACAAAAGTGGGTAGAAATTTACTATACATTTTCCCAATTTTACTATT 3214
 DB 602 CAAAGAAACATTAACAAAAGTGGGTAGAAATTTACTATACATTTTCCCAATTTTACTATT 661
 QY 3215 CAGAAATGCTGTTATTTAGTATGAGATTAAGCACTGATGAAATGCTTTTAAATAC 3274
 DB 662 CAGAAATGCTGTTATTTAGTATGAGATTAAGCACTGATGAAATGCTTTTAAATAC 721
 QY 3275 TATCAGTTAAACATTTATATATGATTAATGATTAATGATTAATGATTAATGATTAAT 3334
 DB 722 TATCAGTTAAACATTTATATATGATTAATGATTAATGATTAATGATTAATGATTAAT 781
 QY 3335 TAAGATCAATTAATAATGATTTTACTCTG 3365
 DB 782 TAAGATCAATTAATAATGATTTTACTCTG 812

RESULT 4
 B0960157 870 bp mRNA lineal EST 21-AUG-2002
 LOCUS AGENCOURT_8949486 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6472264
 DEFINITION 5', mRNA sequence.
 ACCESSION B0960157
 VERSION B0960157.1 GI:22375635
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 870)
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
 Clone distribution by: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: L14M14005 row: 0 column: 17
 High quality sequence stop: 738.
 Location/Qualifiers

FEATURES
 source

1..870
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_id="NIH-MGC-71"
 /library="leiomysarcoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: uterus; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: 01190 dt.
 Average insert size 2.1 kb."
 BASE COUNT 319 a 131 c 204 g 214 t 2 others
 ORIGIN

Query Match 19.4% Score 652; DB 14; Length 870;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 702; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1976 ATTGAGAGAGTACAAATGAGGCGCTTCAAAATTAAGACAAATTCGAATATGATGCGT 2035
 DB 36 ATTGAGAGAGTACAAATGAGGCGCTTCAAAATTAAGACAAATTCGAATATGATGCGT 95
 QY 2036 ATACGATCTTGAAACCTTCTATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2095
 DB 96 ATACGATCTTGAAACCTTCTATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 155
 QY 2096 ATGATAGTATGAT 2155
 DB 156 ATGATAGTATGAT 215
 QY 2156 TAAAGAAACCTTGAAACCTTGATGATGATGATGATGATGATGATGATGATGATGAT 2215
 DB 216 TAAAGAAACCTTGAAACCTTGATGATGATGATGATGATGATGATGATGATGATGAT 2275
 QY 2216 ACAATTAATGTTGAAAGCGTGGCTGAAACCGCAATTAATTAATTAATTAATTAATTA 2275
 DB 276 ACAATTAATGTTGAAAGCGTGGCTGAAACCGCAATTAATTAATTAATTAATTAATTA 335
 QY 2276 AATTAGAATATACATTAATGAGCAATTAATGAGTATGAGTATGAGTATGAGTATGAG 2335
 DB 336 AATTAGAATATACATTAATGAGCAATTAATGAGTATGAGTATGAGTATGAGTATGAG 395
 QY 2336 TCTTTACAAAAACAGACAGAGATGATGATGATGATGATGATGATGATGATGATGATG 2395
 DB 396 TCTTTACAAAAACAGACAGAGATGATGATGATGATGATGATGATGATGATGATGATG 455
 QY 2396 AATTGCTGAGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATG 2455
 DB 456 AATTGCTGAGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATGAGTATG 515
 QY 2456 TCAAAACCATGACAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2515
 DB 516 TCAAAACCATGACAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 575
 QY 2516 TCAATCTGCTATGCTGACAGAGTGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2575
 DB 576 TAAATCTGCTATGCTGACAGAGTGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 635

QY 2576 TTGTTATCCCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2635
 DB 636 TTGTTATCCCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 695
 QY 2636 CCAGAGCTGATGAG 2678
 DB 696 CCAGAGCTGATGAG 738

RESULT 5
 BM476961
 LOCUS BM476961
 DEFINITION AGENCOURT_6481569 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:555523
 ACCESSION 5', mRNA sequence.
 VERSION BM476961.1 GI:18526003
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1046)
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: L14M12275 row: n column: 04
 High quality sequence stop: 622.
 Location/Qualifiers

FEATURES
 source

1..1046
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_id="IMAGE:555523"
 /clone_lib="NIH-MGC-71"
 /library="leiomysarcoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: uterus; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: 01190 dt.
 Average insert size 2.1 kb."
 BASE COUNT 318 a 217 c 227 g 282 t 2 others
 ORIGIN

Query Match 18.4% Score 619; DB 13; Length 1046;
 Best Local Similarity 100.0%; Pred. No. 1; 36-306;
 Matches 619; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 172 TCGAATGGGTATTCACAG 221
 DB 1 TCGAATGGGTATTCACAG 60
 QY 232 GTGAATGTACATCCAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 291
 DB 61 GTGAATGTACATCCAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
 QY 292 GTGAAGAGAGAGATTCAG 351
 DB 121 GTGAAGAGAGAGATTCAG 180
 QY 352 CTGCTGACACCTTGAG 411
 DB 181 CTGCTGACACCTTGAG 240
 QY 412 GCCCTCGGAG 471

Db 241 GCCCTCGGAGAACCGGACGCTCTGGCCGCCGCTACATGACCTGATCTCAGGAG 300
 Oy 472 TTGCCCTCCATCGTTGAGAGCGCTATGATATCTTCACTGCTTAACTCCTT 531
 Db 301 TTGCCCTCCATCGTTGAGAGCGCTATGATATCTTCACTGCTTAACTCCTT 360
 Oy 532 CAGCCCACTGCTGAGACAGCTTGTAGTAGAGAGCTTGTGATAGCTGAGAG 591
 Db 361 CAGCCCACTGCTGAGACAGCTTGTAGTAGAGAGCTTGTGATAGCTGAGAG 420
 Oy 592 GAATGTTGACATTTGAGACAGAAACCGGATTGCTGTCGACAAACAAATGAA 651
 Db 421 GAATGTTGACATTTGAGACAGAAACCGGATTGCTGTCGACAAACAAATGAA 480
 Oy 652 TAGGTTGAGAGAGCTTAAAGAGATGTCGACAAAGAAAGAAAGCTGCTGAT 711
 Db 481 TAGGTTGAGAGAGCTTAAAGAGATGTCGACAAAGAAAGAAAGCTGCTGAT 540
 Oy 712 CTGAATGTTGCTGTCACAAACAGAAACAAATGATGATGCTGACAAAGTAA 771
 Db 541 CTGAATGTTGCTGTCACAAACAGAAACAAATGATGATGCTGACAAAGTAA 600
 Oy 772 TGCTCAGAAAGCAATGAC 790
 Db 601 TGCTCAGAAAGCAATGAC 619

RESULT 6
 LOCUS BF337464 1035 bp mRNA linear EST 22-NOV-2000
 DEFINITION 602035195f1 NIH_CGAP_Brn64 Homo sapiens cDNA clone IMAGE:4183126
 ACCESSION BF337464
 VERSION BF337464.1 GI:11283715
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: David N. Louis, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLM9498 row: n column: 23
 High quality sequence start: 5
 High quality sequence stop: 695.
 Location/Qualifiers
 1. 1035
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4183126"
 /clone_lib="NCI_CGAP_Brn64"
 /tissue_type="glioblastoma with EGFR amplification"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: brain; Vector: pCMV-Sport0; Site: 1; Note: Site: 2; Salt: Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.57 kb. Constructed by Life Technologies. Note: this is a NIH_CGAP library."
 BASE COUNT 353 a 221 c 243 g 218 t
 ORIGIN

Query Match 18.0%; Score 607; DB 2; Length 1035;
 Best Local Similarity 99.8%; Pred. No. 2e-307;
 Matches 657; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1049 AGAATGTGGCAGCAAGACATCCCGGAGCCAGAACTCCAGCTAGAGGCTTAAAGATG 1108
 Db 4 AGAATGTGGCAGCAAGACATCCCGGAGCCAGAACTCCAGCTAGAGGCTTAAAGATG 63
 Oy 1109 AAGTTGCCACCGACCTTGGAGGAGAAATATCATCATCTGCTCCTTAAGAGATG 1168
 Db 64 AAGTTGCCACCGACCTTGGAGGAGAAATATCATCATCTGCTCCTTAAGAGATG 123
 Oy 1169 GAAATCCAGAGTGGCTTTACATTCGCAAGATCATTAGACAAAGCAATAGAT 1228
 Db 124 GAAATCCAGAGTGGCTTTACATTCGCAAGATCATTAGACAAAGCAATAGAT 183
 Oy 1229 CTGAGCCGTGAAAGTTATGCTTGTGCAATTAAGTACTGCTAGTAAATGCTGCT 1288
 Db 184 CTGAGCCGTGAAAGTTATGCTTGTGCAATTAAGTACTGCTAGTAAATGCTGCT 243
 Oy 1289 GCAAGAGTCCACCATTTTGAAGAAATGCTATGCTGATTTGATTAATATGATA 1348
 Db 244 GCAAGAGTCCACCATTTTGAAGAAATGCTATGCTGATTTGATTAATATGATA 303
 Oy 1349 CCGAAGTGAATATCATTTCCAGAGTTGCAAGTCTGTGATTTATTAAGATA 1408
 Db 304 CCGAAGTGAATATCATTTCCAGAGTTGCAAGTCTGTGATTTATTAAGATA 363
 Oy 1409 CTCAAATGCTTGAAGAACTGCTTAACTGGAATGGAAGATGCTGCTGCTCAAT 1468
 Db 364 CTCAAATGCTTGAAGAACTGCTTAACTGGAATGGAAGATGCTGCTGCTCAAT 423
 Oy 1469 TGTGAGCTTTCCCTCATTTATCATTTGATGATGATGATGATGATGATGAT 1528
 Db 424 TGTGAGCTTTCCCTCATTTATCATTTGATGATGATGATGATGATGATGAT 483
 Oy 1529 ATATATCATATGAGGATTTATTTGATGCAAGTGTAAAAAATATGACTAAGAA 1588
 Db 484 ATATATCATATGAGGATTTATTTGATGCAAGTGTAAAAAATATGACTAAGAA 543
 Oy 1589 AAAACAAACCATGATGCTTCCCTGCTGATGATGATGATGATGATGATGAT 1648
 Db 544 AAAACAAACCATGATGCTTCCCTGCTGATGATGATGATGATGATGATGAT 603
 Oy 1649 GAGGGGCGACGAGAGCAAGCAAGCTGAAGAACATTTAAACTATGTAATAT 1706
 Db 604 GAGGGGCGACGAGAGCAAGCAAGCTGAAGAACATTTAAACTATGTAATAT 661

RESULT 7
 LOCUS BF983236 729 bp mRNA linear EST 23-JAN-2001
 DEFINITION 602305873f1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:417083 5',
 mRNA sequence.
 ACCESSION BF983236
 VERSION BF983236.1 GI:12386048
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLM10097 row: a column: 20
 High quality sequence stop: 665.

QY 2946 GCATCAGCTCAATATGACCCGAGATTCCAGAACTTACATCTTAAGCT/AAACAAAGC 3005
 DB 439 GCATCAGCTCAATATGACCCGAGATTCCAGAACTTACATCTTAAGCT/AAACAAAGC 380
 QY 3006 ACTGAAAGAAAGAGTGTCCGACATCAATTAATGTGTAAATATGTGTGAATGTGCCA 3065
 DB 379 ACTGAAAGAAAGAGTGTCCGACATCAATTAATGTGTAAATATGTGTGAATGTGCCA 320
 QY 3066 GCGTGGGAGCAATGATGTGTGCAAGAGGCTTAGATTGCTTTGTGTCAAAATAGGAA 3125
 DB 319 GCGTGGGAGCAATGATGTGTGCAAGAGGCTTAGATTGCTTTGTGTCAAAATAGGAA 260
 QY 3126 TTTTGTAGTGTGTTCAAAATTAATTCACAAAGAACAAATTA/AAAAAGTGGTAGAAT 3195
 DB 259 TTTTGTAGTGTGTTCAAAATTAATTCACAAAGAACAAATTA/AAAAAGTGGTAGAAT 200
 QY 3186 ACCTATGACATTTCCCAATCTTGACATATTCAGATTCGTTATTTAGTCAATGAGATTA 3245
 DB 199 ACCTATGACATTTCCCAATCTTGACATATTCAGATTCGTTATTTAGTCAATGAGATTA 140
 QY 3246 GCATTTGATGACATCTTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3305
 DB 139 GCATTTGATGACATCTTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 80
 QY 3306 AATGATTTCAATTAATGCTACAGAACTGACATTAAGATTAATTAATTAATTAATTAAT 3365
 DB 79 AATGATTTCAATTAATGCTACAGAACTGACATTAAGATTAATTAATTAATTAATTAATTA 20

RESULT 11
 BP686405 781 bp mRNA linear EST 22-DEC-2000
 LOCUS 602143786F1 NIH_MGC_46 Homo sapiens cDNA clone IMAGE:404805 5'
 DEFINITION mRNA sequence.
 ACCESSION BP686405
 VERSION BP686405.1 GI:11971813
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 781)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: rgs@pds-remail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Ling Hong/John's Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LCM1170 row: 1 column: 22
 High quality sequence stop: 731.
 Location/Qualifiers
 1. 781

FEATURES
 source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4304805"
 /clone_id="NIH_MGC_46"
 /tissue_type="leiomyosarcoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: uterus; Vector: pGB7; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned
 into EcoRI/XhoI sites using the following 5' adaptor:
 GGCAAGAGG(5). Size selected >50bp for average insert size
 1.8kb. Library constructed by Ling Hong in the laboratory
 of Gerald M. Rubin (University of California, Berkeley)
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript
 II RT (Life Technologies). Note: this is a NIH_MGC

Library.
 BASE COUNT 235 a 168 c 204 g 174 t
 ORIGIN
 Query Match 14.5%; Score 489; DB 12; Length 791.
 Best Local Similarity 99.8%; Pred. No. 1e-239;
 Matches 609; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 392 GGACTGGGAATTCGTGGAGGCGCTCGGAGAACCGGACCGCTGTGGCGTTCCTATTA 451
 DB 12 GGACTGGGAATTCGTGGAGGCGCTCGGAGAACCGGACCGCTGTGGCGTTCCTATTA 71
 QY 452 TGAACCTGAGCTCAGGAGCTTGCCTCTCCATGTTTGAAGAGCTCATTAATATC 511
 DB 72 TGAACCTGAGCTCAGGAGCTTGCCTCTCCATGTTTGAAGAGCTCATTAATATC 131
 QY 512 TCCAGCTGCGAAGCTCCTCCAGCCGACCTGGTGGCAAGCTTGTAGTATTAAGCT 571
 DB 132 TCCAGCTGCGAAGCTCCTCCAGCCGACCTGGTGGCAAGCTTGTAGTATTAAGCT 190
 QY 572 TGGATTAAGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 631
 DB 191 TGGATTAAGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 250
 QY 632 CAGAAACAAATGGAATGAATCAATCAAGTGTAAAGAGCTACTAAAGAGATTG 691
 DB 251 CAGAAACAAATGGAATGAATCAATCAAGTGTAAAGAGCTACTAAAGAGATTG 310
 QY 692 AAAAGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 751
 DB 311 AAAAGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 370
 QY 752 AAGAGTTACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 811
 DB 371 AAGAGTTACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 430
 QY 812 TTGATGCTCTCAAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 871
 DB 431 TTGATGCTCTCAAGTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 490
 QY 872 AGGAGGCTGGGCGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 931
 DB 491 AGGAGGCTGGGCGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 550
 QY 932 TTTCAGATGACAGACAAAGTTTGGCAGAGAGGAGTGCAGCTGTTAGATGAAGTTTG 991
 DB 551 TTTCAGATGACAGACAAAGTTTGGCAGAGAGGAGTGCAGCTGTTAGATGAAGTTTG 610
 QY 992 GACATACAG 1001
 DB 611 GACATACAG 620

RESULT 12
 A1718277/c 536 bp mRNA linear EST 10-JUN-1999
 LOCUS A1718277
 DEFINITION as51e06.x1 Barstead aorta HPLK6 Homo sapiens cDNA clone
 IMAGE:2320738 3', mRNA sequence.
 ACCESSION A1718277
 VERSION A1718277.1 GI:5035533
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 536)
 Reference
 Hillier, L., Allen, M., Bowles, L., Dubuque, T., Gelfand, J., Jost, S.,
 Krizman, D., Kudaba, T., Lacy, M., Le, N., Lemon, G., Mirta, M., Martin,
 J., Moore, B., Schellenberg, K., Steptoe, M., Tan, J., Welling, B.,
 White, Y., Wylie, T., Waterston, R., and Wilson, R.
 WashU-NCI human EST Project
 Unpublished (1997)
 Contact: Wilson RK

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu

This clone is available royalty-free through LNL; contact the
IMAGE Consortium (infoimage.lnl.gov) for further information.
Seq primer: -400P from Glibco
High quality sequence stop: 481.

FEATURES

source
1. 536
/organism="Homo sapiens"
/db.xref="taxon:9606"
/clone="IMAGE:2320738"
/clone.lib="Barstead aorta HPLRB6"
/sex="male"
/dev_stage="adult, age 64"
/lab_host="DH10B (phage resistant)"
/note="Organ: aorta; Vector: p173D-Pac (Pharmacia) with a
modified polylinker; Site:1: EcoRI; Site:2: NotI; 1st
strand cDNA was primed with a Not I - c190(dT) primer (5'
37); double-stranded cDNA was ligated to Eco RI adaptors
(5' AATTGATCGAC 3' and 5' GTTGGATCG 37), digested
with Not I and cloned into the Not I and Eco RI sites of
the modified pT73 vector. Library constructed by Bob
Barstead."

BASE COUNT 157 a 93 c 84 g 202 t
ORIGIN

Query Match 14.3%; Score 482; DB 9; Length 716;
Best Local Similarity 99.8%; Pred. No. 3.8e-216;
Matches 532; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2833 AAGAAATATTTGGCAACATTACAGAAATACCATCATATTAATCTTCTTTGGCAA 2892
|||||
DB 536 AAGAAATATTTGGCAACATTACAGAAATACCATCATATTAATCTTCTTTGGCAA 477
OY 2893 AACTGAGTGTGCTAGCTGTCTGGGAAAGATACCATGTAAATGAGAAATGATCAG 2952
|||||
DB 476 AACTGAGTGTGCTAGCTGTCTGGGAAAGATACCATGTAAATGAGAAATGATCAG 417
OY 2953 GTCAATATGACCCGAGAAATTCAGAACTTACATTTGAAGAAAGAAAGCACTGCA 3012
|||||
DB 416 GTCAATATGACCCGAGAAATTCAGAACTTACATTTGAAGAAAGAAAGCACTGCA 357
OY 3013 AAGAAGTGTGCGACTATTCATAAATGTAATCATCTGTAATGTGCTGAGCTTGG 3072
|||||
DB 356 AAGAAGTGTGCGACTATTCATAAATGTAATCATCTGTAATGTGCTGAGCTTGG 297
OY 3073 GGAACAATGATGTGCGACAAAGGCTTAGATTGCTTGGTAAATATGATTTGTA 3132
|||||
DB 296 GGAACAATGATGTGCGACAAAGGCTTAGATTGCTTGGTAAATATGATTTGTA 237
OY 3133 GTGCTTTCAAAAATTAATTCACAAAGAAACAAATACAAAATGAGTAATTAACCTA 3192
|||||
DB 236 GTGCTTTCAAAAATTAATTCACAAAGAAACAAATACAAAATGAGTAATTAACCTA 177
OY 3193 ACATTTCGAATCTTGACTATTCAGAAATGCTGTTATTATTAGTATGAGAGG TAGACTTG 3252
|||||
DB 176 ACATTTCGAATCTTGACTATTCAGAAATGCTGTTATTATTAGTATGAGAGG TAGACTTG 117
OY 3253 ATGGAAGATCTTTTAAATATCATCATGTTAAATTAATATGATTAATTAATGAT 3312
|||||
DB 116 ATGGAAGATCTTTTAAATATCATCATGTTAAATTAATATGATTAATTAATGAT 57
OY 3313 TCATTATGCTACAGAACTGACATAGAAATCAATAAATGATTTTATCTG 3365
|||||
DB 56 TCATTATGCTACAGAACTGACATAGAAATCAATAAATGATTTTATCTG 4

RESULT 13
AM589567/c

LOCUS AM589567 475 bp mRNA linear ES: 22-MAR-2000
DEFINITION x012b12.x1 NCI CGAP UT3 Homo sapiens cDNA clone IMAGE:2320738 3'
similar to TR:095786 095786 RNA HELICASE.;, mRNA sequence.
ACCESSION AM589567
VERSION AM589567.1 GI:7276681
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo;
REFERENCE 1 (bases 1 to 475)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9aput-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
image.lnl.gov/image/html/lresources.shtml

FEATURES

source
1. 475
/organism="Homo sapiens"
/db.xref="taxon:9606"
/clone="IMAGE:2703743"
/tissue_type="poorly-differentiated endometrial
adenocarcinoma, 2 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pCMV-SPORT1; Site:1: SalI;
Site:2: NotI. Cloned unidirectionally. Primer: oligo dT.
Average insert size 1.45 kb. Life Technologies catalog #:
11541-018"

possible reversed clone: polyT not found
Seq primer: -400P from Glibco
High quality sequence stop: 407.

BASE COUNT

142 a 88 c 71 g 174 t
ORIGIN

Query Match 14.1%; Score 473; DB 10; Length 475;
Best Local Similarity 100.0%; Pred. No. 1.6e-231;
Matches 473; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2893 AACTGAGTGTGCTAGCTGTCTGGGAAAGATATCATGTAATTGAGAAATATTAAC 2952
|||||
DB 475 AACTGAGTGTGCTAGCTGTCTGGGAAAGATATCATGTAATTGAGAAATATTAAC 416
OY 2953 GTCAATATGACCCGAGAAATTCAGAACTTACATTTGAAGAAAGAAACAAAT 3012
|||||
DB 415 GTCAATATGACCCGAGAAATTCAGAACTTACATTTGAAGAAAGAAACAAAT 356
OY 3013 AAGAAGTGTGCGACTATTCATAAATGTAATCATCTGTAATGTGCTGAGCTTGG 3072
|||||
DB 355 AAGAAGTGTGCGACTATTCATAAATGTAATCATCTGTAATGTGCTGAGCTTGG 296
OY 3073 GGAACAATGATGTGCGACAAAGGCTTAGATTGCTTGGTAAATATGATTTGTA 3132
|||||
DB 295 GGAACAATGATGTGCGACAAAGGCTTAGATTGCTTGGTAAATATGATTTGTA 236
OY 3133 GTGCTTTCAAAAATTAATTCACAAAGAAACAAATACAAAATGAGTAATTAACCTA 3192
|||||
DB 235 GTGCTTTCAAAAATTAATTCACAAAGAAACAAATACAAAATGAGTAATTAACCTA 176
OY 3193 ACATTTCGAATCTTGACTATTCAGAAATGCTGTTATTATTAGTATGAGAGTAGACTTG 3252
|||||
DB 175 ACATTTCGAATCTTGACTATTCAGAAATGCTGTTATTATTAGTATGAGAGTAGACTTG 116

```

OY 3253 ATGGAAGATCTTTTAAATCTATCGATTAACATTTAATATCTATCTATATGAT 3312
DB 115 ATGGAAGATCTTTTAAATCTATCGATTAACATTTAATATCTATCTATATGAT 56
OY 3313 TCATTATCTGACAGACTGACATTAAGATCAATTAATGATGATGATGATG 3365
DB 55 TCATTATCTGACAGACTGACATTAAGATCAATTAATGATGATGATGATG 3

RESULT 14
A1922705 521 bp mRNA linear EST 01-SEP-1999
LOCUS w011f08.k1 NCI_CGAP_Pan1 Homo sapiens cDNA clone IMAGE:2455047 3'
DEFINITION mRNA sequence.
ACCESSION A1922705
VERSION A1922705.1 GI:5658669
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 521)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Life Technologies catalog #: 11548-013
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distributor information can be
found through the I.M.A.G.E. Consortium/BLNI at:
www.bio.litl.gov/bdrp/image/image.html
Seq primer: -400p from GIBCO
High quality sequence stop: 410.
FEATURES
source
1..521
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="NCI_CGAP_Pan1"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/note="Organ: pancreas; Vector: pCMV-Sirt6; Site_1: Salt;
Site_2: NotI; Cloned unidirectionally; Primer: oligo dt.
Average insert size 1.72 kb. Life Technologies catalog #:
11548-013"
BASE COUNT 154 a 90 c 82 g 195 t
ORIGIN
Query Match 13.8%; Score 466; DB 9; Length 521;
Best Local Similarity 99.8%; Pred. No. 6,4e-228;
Matches 516; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2849 AGCATTACAGAAATACCATCAGTAATACCTTCTTCCAAAAGTCAGTGTAG 2908
DB 521 AGCATTACAGAAATACCATCAGTAATACCTTCTTCCAAAAGTCAGTGTAG 462
OY 2909 CCTGTTGGGAGAGATATCATGTAATGAGAAATGCATCATCTCAATATGACCCAG 2968
DB 461 CCTGTTGGGAGAGATATCATGTAATGAGAAATGCATCATCTCAATATGACCCAG 402
OY 2969 AATTCAAGAACTTTACATGTAAGAGAAAACAAAGCACAGTAAAGAAAGTGTGCCACT 3028
DB 401 AATTCAAGAACTTTACATGTAAGAGAAAACAAAGCACAGTAAAGAAAGTGTGCCACT 342
OY 3029 ATCAATATAATGTAATCATCTGCAATGAGGAGAGCTTACAGAAATGATGATGTC 3088
DB 341 ATCAATATAATGTAATCATCTGCAATGAGGAGAGCTTACAGAAATGATGATGTC 282
OY 3089 ACAAAGCTTACATTTGCTGCTCAAAATTAAGGAATTTTCAAGTGTTCACAAATA 3148
DB 281 ACAAAGCTTACATTTGCTGCTCAAAATTAAGGAATTTTCAAGTGTTCACAAATA 222

```

```

OY 3149 ATTCACAAAGAAACATACACAAAGAGGCTAGCAATTTACTTACATTCCTAAATTTG 3208
DB 221 ATTCACAAAGAAACATACACAAAGAGGCTAGCAATTTACTTACATTCCTAAATTTG 162
OY 3209 ACTATTGAGAACTGCTGTTTATTAGTAGAGAGATAGACCTGATGAAACATTTTCA 3268
DB 161 ACTATTGAGAACTGCTGTTTATTAGTAGAGATAGACCTGATGAAACATTTTCA 102
OY 3269 AAATACTATCAGTTAAACATTTAATGATTATGATTATGATTATTTACACAA 3328
DB 101 AAATACTATCAGTTAAACATTTAATGATTATGATTATGATTATTTACACAA 42
OY 3329 CTGACATTAAGAAATCAATTAATGATTTTACTCTG 3365
DB 41 CTGACATTAAGAAATCAATTAATGATTTTACTCTG 5

RESULT 15
BE940626 546 bp mRNA linear EST 02-OCT-2000
LOCUS BE940626
DEFINITION RC3-UT0064-170800-021-e06 UT0064 Homo sapiens cDNA mRNA sequence.
ACCESSION BE940626
VERSION BE940626.1 GI:10470130
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 546)
AUTHORS Dias Neto E., Garcia Correa R., Verjovsky-Almeida S., Hirones M.R.,
Negal M.A., da Silva M. Jr., Zago M.A., Bordin S., Costa F.F.,
Goldman G.H., Carvalho A.F., Matsukuma A., Bala G.C., Simpson D.H.,
Brunstein A., de Oliveira P.S., Bucher P., Jongeneel V., O'Hare
M.J., Soares F., Brentani R.R., Reis L.F., de Souza S.J. and
Simpson A.J.
TITLE Shotgun sequencing of the human transcriptome with 3' expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 2020263
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL:
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=2_R3-UT0064-170
800-021-e06&t3=2000-08-17&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 53
High quality sequence stop: 542.
FEATURES
source
1..546
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="UT0064"
/dev_stage="Adult"
/note="Organ: uterus-tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 205 a 86 c 102 g 153 t
ORIGIN
Query Match 13.8%; Score 463; DB 12; Length 546;
Best Local Similarity 100.0%; Pred. No. 2.3e-226;

```

Matches 463: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 2844 TGGCAGAGATTACAGATTAACCCATCTATTAACCTTCTTGGCAAAATCTGCAGTGT 2903
      |||||||
Db 77 TGGCAGAGATTACAGATTAACCCATCTATTAACCTTCTTGGCAAAATCTGCAGTGT 136
      |||||||
OY 2904 GCTACCTGTTCTGGGAGATATCATGTATGAGAAATGCAATGCACTGCATATGAC 2963
      |||||||
Db 137 GCTACCTGTTCTGGGAGATATCATGTATGAGAAATGCAATGCACTGCATATGAC 196
      |||||||
OY 2964 CCCAGATTCAGGAACCTTACATTTGTAAGAGAAACCAAGCACTGCAGAAAGAGTGTGC 3023
      |||||||
Db 197 CCCAGATTCAGGAACCTTACATTTGTAAGAGAAACCAAGCACTGCAGAAAGAGTGTGC 256
      |||||||
OY 3024 CGACTATCAATTAATGTTGAATCATCTGCAAAATGTGGCAGGCTTGGCAACATGAT 3083
      |||||||
Db 257 CGACTATCAATTAATGTTGAATCATCTGCAAAATGTGGCAGGCTTGGCAACATGAT 316
      |||||||
OY 3084 GGTGCACAAAGGCTAGATTGGCTTGTCTCAAAATAGAAATTTTGTAGTGGTTTCAA 3143
      |||||||
Db 317 GGTGCACAAAGGCTAGATTGGCTTGTCTCAAAATAGAAATTTTGTAGTGGTTTCAA 376
      |||||||
OY 3144 AATTAATTCACAAAGAAACATATACAAAAGTGGGTAGAAATACCTATCAATTTCCCA 3203
      |||||||
Db 377 AATTAATTCACAAAGAAACATATACAAAAGTGGGTAGAAATACCTATCAATTTCCCA 436
      |||||||
OY 3204 TCTTGACTATTCAGAAATGCTTTTATTAGTAGAGATTAAGCACTTGATTTGAAGATTC 3263
      |||||||
Db 437 TCTTGACTATTCAGAAATGCTTTTATTAGTAGAGATTAAGCACTTGATTTGAAGATTC 496
      |||||||
OY 3264 TTTTAAATACATCATGTTAAACATTTAATAGATTATGATTA 3306
      |||||||
Db 497 TTTTAAATACATCATGTTAAACATTTAATAGATTATGATTA 539
      |||||||
```

Search completed: May 17, 2003, 11:01:43
Job time : 4534 secs

1592 ACAACCCAGTGTCCCTTCCTCAGATACTGGGACTAACAGCTTCACCTGGTGTGGAG 1651

|||||
Db 1 AACAAACAGTATCCCTTCCTCAGATACGTGGAGCTAACAGTTCACCTTCCTGTTGAG 60
QY 1652 GGCCGACGAGCAAGCAAGTGAAGACATTTAAACATGTC /MCTTGAG 1711
|||||
Db 61 GGCCGACGAGCAAGCAAGTGAAGACATTTAAACATGTC /MCTTGAG 120
QY 1712 CATTTACTTAACTGTTAAAGAAACCTTGATCACTGAAACCA /ACAGGAGC 1771
|||||
Db 121 CATTTACTTAACTGTTAAAGAAACCTTGATCACTGAAACCA /ACAGGAGC 180
|||||
RESULT 2
US-08-143-576-6
: Sequence 6, Application US/08143576
: Patent No. 5643761
: GENERAL INFORMATION:
: APPLICANT: Jiaqiang, Paul B.
: TITLE OF INVENTION: METHOD FOR GENERATING A SUBSTRUCTURE
: NUMBER OF SEQUENCES: 16
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: John P. White, C/O Cooper & Dunham
: STREET: 30 Rockefeller Plaza
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10112
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/143,576
: FILING DATE: 25-OCT-1993
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: White, John P.
: REGISTRATION NUMBER: 28,678
: REFERENCE/DOCKET NUMBER: 43563/JPW/AKC
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 977-9550
: TELEFAX: (212) 664-0525
: TELEX: 422523 COOP UI
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 193 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
US-08-143-576-6
Query Match 3.2%; Score 108; DB 1; Length 193;
Best Local Similarity 100.0%; Ptd. No. 3.3e-41;
Matches 108; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Patent No. 6444465
: GENERAL INFORMATION:
: APPLICANT: C. Frank Bennett
: APPLICANT: Jacqueline Freier
: TITLE OF INVENTION: OLIGONUCLEOTIDE INHIBITION OF HER-1 EXPRESSION
: FILE REFERENCE: R15-0138
: CURRENT FILING DATE: US/09/676,610B
: NUMBER OF SEQ ID NOS: 182
: SEQ ID NO 24
: LENGTH: 169998
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: exon
: LOCATION: (1208)...(1472)
: NAME/KEY: intron
: LOCATION: (1473)...(124390)
: NAME/KEY: exon
: LOCATION: (124391)...(124544)
: NAME/KEY: intron
: LOCATION: (124545)...(125409)
: NAME/KEY: exon
: LOCATION: (125410)...(125595)
: NAME/KEY: intron
: LOCATION: (125596)...(128711)
: NAME/KEY: exon
: LOCATION: (128712)...(128848)
: NAME/KEY: intron
: LOCATION: (128849)...(133400)
: NAME/KEY: exon
: LOCATION: (133401)...(133469)
: NAME/KEY: intron
: LOCATION: (133470)...(134652)
: NAME/KEY: exon
: LOCATION: (134653)...(134773)
: NAME/KEY: intron
: LOCATION: (134774)...(136116)
: NAME/KEY: exon
: LOCATION: (136117)...(136261)
: NAME/KEY: intron
: LOCATION: (136262)...(137936)
: NAME/KEY: exon
: LOCATION: (137937)...(138053)
: NAME/KEY: intron
: LOCATION: (138054)...(138637)
: NAME/KEY: exon
: LOCATION: (138638)...(138766)
: NAME/KEY: intron
: LOCATION: (138767)...(138864)
: NAME/KEY: exon
: LOCATION: (138865)...(138940)
: NAME/KEY: intron
: LOCATION: (138941)...(139765)
: NAME/KEY: exon
: LOCATION: (139766)...(139860)
: NAME/KEY: intron
: LOCATION: (139861)...(142245)
: NAME/KEY: exon
: LOCATION: (142246)...(142445)
: NAME/KEY: intron
: LOCATION: (142446)...(143605)
: NAME/KEY: exon
: LOCATION: (143606)...(143738)
: NAME/KEY: intron
: LOCATION: (143739)...(145838)
: NAME/KEY: exon
: LOCATION: (145839)...(145931)
: NAME/KEY: intron
: LOCATION: (145932)...(147385)
: NAME/KEY: exon
: LOCATION: (147386)...(147544)


```

NAME/KEY: Intron
LOCATION: (147545)...(151274)
NAME/KEY: exon
LOCATION: (153275)...(153321)
NAME/KEY: Intron
LOCATION: (153322)...(155088)
NAME/KEY: exon
LOCATION: (155089)...(155231)
NAME/KEY: Intron
LOCATION: (155232)...(156025)
NAME/KEY: exon
LOCATION: (156026)...(156151)
NAME/KEY: Intron
LOCATION: (156152)...(156826)
NAME/KEY: exon
LOCATION: (156827)...(156928)
NAME/KEY: Intron
LOCATION: (156929)...(163399)
NAME/KEY: exon
LOCATION: (163400)...(163586)
US-09-676-610B-24

```

```

Query Match
Best Local Similarity 100.0%; Score 22; DB 4; Length 14998;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 2102 GTGATGAGGTGTGTGATGA 2123
Db 149077 GTGATGAGGTGTGTGATGA 149098

```

```

RESULT 4
US-09-345-882-12
Sequence 12, Application US/09345882
Patent No. 6399373
GENERAL INFORMATION:
APPLICANT: Bouquellet, Lydie
TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOLASTOMA B-INDING PROTEIN (RBP-7)
FILE REFERENCE: GENSET 031A
CURRENT APPLICATION NUMBER: US/09/345,882
CURRENT FILING DATE: 1999-06-30
PRIOR APPLICATION NUMBER: US 60/091,315
PRIOR FILING DATE: 1998-06-30
PRIOR APPLICATION NUMBER: US 60/111,909
PRIOR FILING DATE: 1998-12-10
NUMBER OF SEQ ID NOS: 140
SOFTWARE: Patent.pm
SEQ ID NO 12
LENGTH: 139
TYPE: DNA
ORGANISM: Homo sapiens
US-09-345-882-12

```

```

Query Match
Best Local Similarity 100.0%; Score 20; DB 4; Length 1418;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 2135 GTGATGAAGATGAGATGAT 2154
Db 30 GTGATGAAGATGAGATGAT 49

```

```

RESULT 5
US-09-111-470-7
Sequence 7, Application US/09111470
Patent No. 6277959
GENERAL INFORMATION:
APPLICANT: Valladeau, Jenny
APPLICANT: Ravel, Odile
APPLICANT: Bates, Elizabeth E.M.
APPLICANT: Ford, John
APPLICANT: Saeland, Sem

```

```

APPLICANT: Lebecque, Serge J.E.
TITLE OF INVENTION: Mammalian Membrane Protein Genes;
TITLE OF INVENTION: Related Reagents
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESS: DNAX Research Institute
STREET: 901 California Avenue
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/111,470
FILING DATE: 08-JUL-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/053,080
FILING DATE: 09-JUL-1997
ATTORNEY/AGENT INFORMATION:
NAME: Ching, Edwin P.
REGISTRATION NUMBER: 34,090
REFERENCE/DOCKET NUMBER: SP0695
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650)852-9196
TELEFAX: (650)496-1200
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1418 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 279..992
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1348
OTHER INFORMATION: /note="poly-A addition motif"
US-09-111-470-7

```

```

Query Match
Best Local Similarity 100.0%; Score 20; DB 4; Length 1418;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1924 GAAAAAAGCTGCAAAAAA 1943
Db 510 GAAAAAAGCTGCAAAAAA 529

```

```

RESULT 6
US-09-345-882-4
Sequence 4, Application US/09345882
Patent No. 6399373
GENERAL INFORMATION:
APPLICANT: Bouquellet, Lydie
TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOLASTOMA BINDING PROTEIN (RBP-7)
FILE REFERENCE: GENSET 031A
CURRENT APPLICATION NUMBER: US/09/345,882
CURRENT FILING DATE: 1999-06-30
PRIOR APPLICATION NUMBER: US 60/091,315
PRIOR FILING DATE: 1998-06-30
PRIOR APPLICATION NUMBER: US 60/111,909
PRIOR FILING DATE: 1998-12-10
NUMBER OF SEQ ID NOS: 140
SOFTWARE: Patent.pm
SEQ ID NO 4

```

```

LENGTH 6002
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
  NAME/KEY: allele
  LOCATION: 1319
  OTHER INFORMATION: 5-130-257 : polymorphic base A G
FEATURE:
  NAME/KEY: allele
  LOCATION: 1338
  OTHER INFORMATION: 5-130-276 : polymorphic base A G
FEATURE:
  NAME/KEY: allele
  LOCATION: 1944
  OTHER INFORMATION: 5-136-174 : polymorphic base C T
FEATURE:
  NAME/KEY: allele
  LOCATION: 3329
  OTHER INFORMATION: 5-143-84 : polymorphic base A G
FEATURE:
  NAME/KEY: allele
  LOCATION: 3346
  OTHER INFORMATION: 5-143-101 : polymorphic base A C
FEATURE:
  NAME/KEY: allele
  LOCATION: 4582
  OTHER INFORMATION: 5-148-352 : polymorphic base G T
FEATURE:
  NAME/KEY: allele
  LOCATION: 1107..1125
  OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID# 1
FEATURE:
  NAME/KEY: allele
  LOCATION: 1107..1125
  OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID# 1
FEATURE:
  NAME/KEY: allele
  LOCATION: 1315..1338
  OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID# 5
FEATURE:
  NAME/KEY: allele
  LOCATION: 1315..1338
  OTHER INFORMATION: polymorphic fragment 5-130-276 SEQ ID# 5
FEATURE:
  NAME/KEY: allele
  LOCATION: 1921..1967
  OTHER INFORMATION: polymorphic fragment 5-136-174 SEQ ID# 1
FEATURE:
  NAME/KEY: allele
  LOCATION: 1921..1967
  OTHER INFORMATION: polymorphic fragment 5-136-174 SEQ ID# 1
FEATURE:
  NAME/KEY: allele
  LOCATION: 3306..3352
  OTHER INFORMATION: polymorphic fragment 5-143-84 SEQ ID# 7
FEATURE:
  NAME/KEY: allele
  LOCATION: 3306..3352
  OTHER INFORMATION: polymorphic fragment 5-143-84 SEQ ID# 7
FEATURE:
  NAME/KEY: allele
  LOCATION: 1296..1338
  OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID# 4
FEATURE:
  NAME/KEY: allele
  LOCATION: 1296..1338
  OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID# 4
FEATURE:
  NAME/KEY: allele
  LOCATION: 3323..3369
  OTHER INFORMATION: polymorphic fragment 5-143-101 SEQ ID# 5
FEATURE:
  NAME/KEY: allele
  LOCATION: 3323..3369
  OTHER INFORMATION: polymorphic fragment 5-143-101 SEQ ID# 5

```

LOCATION: 3323..3369	OTHER INFORMATION: polymorphic fragment 5-143-101 SEQ ID#6
FEATURE:	
NAME/KEY: allele	
LOCATION: 4559..4605	
OTHER INFORMATION: polymorphic fragment 5-148-352 SEQ ID#4	
FEATURE:	
NAME/KEY: allele	
LOCATION: 4559..4605	
OTHER INFORMATION: polymorphic fragment 5-148-352 SEQ ID#6	
FEATURE:	
NAME/KEY: misc.feature	
LOCATION: 442..444	
OTHER INFORMATION: ATG	
FEATURE:	
NAME/KEY: misc.feature	
LOCATION: 4378..4380	
OTHER INFORMATION: stop : TGA	
FEATURE:	
NAME/KEY: polyA_signal	
LOCATION: 4878..4883	
OTHER INFORMATION: potential	
FEATURE:	
NAME/KEY: polyA_signal	
LOCATION: 5116..5121	
OTHER INFORMATION: potential	
FEATURE:	
NAME/KEY: polyA_signal	
LOCATION: 5896..5901	
OTHER INFORMATION: potential	
FEATURE:	
NAME/KEY: polyA_signal	
LOCATION: 5981..5986	
FEATURE:	
NAME/KEY: misc.feature	
LOCATION: 209..756	
OTHER INFORMATION: homology with EST in ref emb1:W645.1	
FEATURE:	
NAME/KEY: misc.feature	
LOCATION: 391..815	
OTHER INFORMATION: complement homology with EST in ref emb1:W37603	
FEATURE:	
NAME/KEY: misc.feature	
LOCATION: 453..898	
OTHER INFORMATION: complement homology with EST in ref emb1:H39516	
FEATURE:	
NAME/KEY: misc.feature	
LOCATION: 818..1306	
OTHER INFORMATION: complement homology with EST in ref emb1:W67770	
FEATURE:	
NAME/KEY: misc.feature	
LOCATION: 844..1303	
OTHER INFORMATION: complement homology with EST in ref emb1:AA262427	
FEATURE:	
NAME/KEY: misc.feature	
LOCATION: 1351..1702	
OTHER INFORMATION: complement homology with EST in ref emb1:AA485189	
FEATURE:	
NAME/KEY: misc.feature	
LOCATION: 1866..2109	
OTHER INFORMATION: homology with EST in ref emb1:AA296993	
FEATURE:	
NAME/KEY: misc.feature	
LOCATION: 2181..2281	
OTHER INFORMATION: homology with EST in ref emb1:761718	
FEATURE:	
NAME/KEY: misc.feature	
LOCATION: 2253..2482	
OTHER INFORMATION: homology with EST in ref emb1:AA082927	
FEATURE:	
NAME/KEY: misc.feature	
LOCATION: 2480..2842	
OTHER INFORMATION: complement homology with EST in ref emb1:H38607	

FEATURE:
NAME/KEY: misc_feature
LOCATION: 3334..3733
OTHER INFORMATION: homology with EST in refembl:A:79595
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3631..3870
OTHER INFORMATION: complement homology with EST in refembl:AL169631
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3683..4221
OTHER INFORMATION: homology with EST in refembl:H:12
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4277..4796
OTHER INFORMATION: homology with EST in refembl:A:79016
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4516..5016
OTHER INFORMATION: homology with EST in refembl:A:79433
FEATURE:
NAME/KEY: misc_feature
LOCATION: 5580..6002
OTHER INFORMATION: complement homology with EST in refembl:AL167428
US-09-345-882-4

Query Match 0.6% Score 20; DB 4; Length 112;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2135 GTGATGAGATGAGATGAT 2154
|||||
DB 917 GTGATGAGATGAGATGAT 916

RESULT 7
US-09-345-882-1
Sequence 1, Application US/09345882
Patent No. 6399373
GENERAL INFORMATION:
APPLICANT: Bouquelere, Lydie
TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOL-BLASHIN A BINDING PROTEIN (RBP-7)
FILE REFERENCE: GENSET 0314
CURRENT APPLICATION NUMBER: US/09/345,882
CURRENT FILING DATE: 1999-06-30
PRIOR APPLICATION NUMBER: US 60/091,315
PRIOR FILING DATE: 1998-06-30
PRIOR APPLICATION NUMBER: US 60/111,909
PRIOR FILING DATE: 1998-12-10
NUMBER OF SEQ ID NOS: 140
SOFTWARE: Patent.pm
SEQ ID NO 1
LENGTH: 162450
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: allele
LOCATION: 72794
OTHER INFORMATION: 5-124-273 : polymorphic base A : G
FEATURE:
NAME/KEY: allele
LOCATION: 88073
OTHER INFORMATION: 5-127-261 : polymorphic base A : C
FEATURE:
NAME/KEY: allele
LOCATION: 90842
OTHER INFORMATION: 99-1437-325 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 93714
OTHER INFORMATION: 5-128-60 : polymorphic base deletion of GT
FEATURE:

NAME/KEY: allele
LOCATION: 97122
OTHER INFORMATION: 99-1442-224 : polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 97152
OTHER INFORMATION: 5-129-144 : polymorphic base deletion of T
FEATURE:
NAME/KEY: allele
LOCATION: 99038
OTHER INFORMATION: 5-130-257 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 99117
OTHER INFORMATION: 5-130-276 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 103806
OTHER INFORMATION: 5-131-395 : polymorphic base A or T
FEATURE:
NAME/KEY: allele
LOCATION: 106940
OTHER INFORMATION: 5-133-375 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 108106
OTHER INFORMATION: 5-135-155 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 108149
OTHER INFORMATION: 5-135-198 : polymorphic base insertion of GTT
FEATURE:
NAME/KEY: allele
LOCATION: 108308
OTHER INFORMATION: 5-135-357 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 108471
OTHER INFORMATION: 5-136-174 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 134134
OTHER INFORMATION: 5-140-120 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 134362
OTHER INFORMATION: 5-140-348 : polymorphic base insertion of A
FEATURE:
NAME/KEY: allele
LOCATION: 134374
OTHER INFORMATION: 5-140-361 : polymorphic base insertion of CA
FEATURE:
NAME/KEY: allele
LOCATION: 146328
OTHER INFORMATION: 5-143-84 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 146345
OTHER INFORMATION: 5-143-101 : polymorphic base A or C
FEATURE:
NAME/KEY: allele
LOCATION: 150329
OTHER INFORMATION: 5-145-24 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 160031
OTHER INFORMATION: 5-148-352 : polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 72771..72817
OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID30
FEATURE:
NAME/KEY: allele

LOCATION: 72771..72817
 OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID# 1
 FEATURE:
 NAME/KEY: allele
 LOCATION: 88050..88096
 OTHER INFORMATION: polymorphic fragment 5-127-271 SEQ ID# 1
 FEATURE:
 NAME/KEY: allele
 LOCATION: 88050..88096
 OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID# 2
 FEATURE:
 NAME/KEY: allele
 LOCATION: 90819..90865
 OTHER INFORMATION: complement polymorphic fragment 99-147-325 SEQ ID# 9
 FEATURE:
 NAME/KEY: allele
 LOCATION: 93690..93736
 OTHER INFORMATION: complement polymorphic fragment 99-147-325 SEQ ID# 7
 FEATURE:
 NAME/KEY: allele
 LOCATION: 93690..93736
 OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID# 7
 FEATURE:
 NAME/KEY: allele
 LOCATION: 97099..97145
 OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID# 7050
 FEATURE:
 NAME/KEY: allele
 LOCATION: 97099..97145
 OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID# 71
 FEATURE:
 NAME/KEY: allele
 LOCATION: 97130..97177
 OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID# 4
 FEATURE:
 NAME/KEY: allele
 LOCATION: 97130..97177
 OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID# 4
 FEATURE:
 NAME/KEY: allele
 LOCATION: 99075..99121
 OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID# 4
 FEATURE:
 NAME/KEY: allele
 LOCATION: 99075..99121
 OTHER INFORMATION: polymorphic fragment 5-130-257 SEQ ID# 4
 FEATURE:
 NAME/KEY: allele
 LOCATION: 99094..99140
 OTHER INFORMATION: polymorphic fragment 5-130-275 SEQ ID# 5
 FEATURE:
 NAME/KEY: allele
 LOCATION: 99094..99140
 OTHER INFORMATION: polymorphic fragment 5-130-275 SEQ ID# 5
 FEATURE:
 NAME/KEY: allele
 LOCATION: 103783..103828
 OTHER INFORMATION: polymorphic fragment 5-131-305 SEQ ID# 6
 FEATURE:
 NAME/KEY: allele
 LOCATION: 103783..103828
 OTHER INFORMATION: polymorphic fragment 5-131-305 SEQ ID# 6
 FEATURE:
 NAME/KEY: allele
 LOCATION: 106918..106966
 OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID# 7
 FEATURE:
 NAME/KEY: allele
 LOCATION: 106918..106966

OTHER INFORMATION: polymorphic fragment 5-133-375 SEQ ID# 8
 FEATURE:
 NAME/KEY: allele
 LOCATION: 108084..108130
 OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID# 8
 FEATURE:
 NAME/KEY: allele
 LOCATION: 108084..108130
 OTHER INFORMATION: polymorphic fragment 5-135-155 SEQ ID# 9
 FEATURE:
 NAME/KEY: allele
 LOCATION: 108127..108177
 OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID# 9
 FEATURE:
 NAME/KEY: allele
 LOCATION: 108127..108177
 OTHER INFORMATION: polymorphic fragment 5-135-198 SEQ ID# 60
 FEATURE:

Query Match 0.68; Score 20; DB 4; Length 162450;
 Best Local Similarity 100.0%; Pred. No. 7.4;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 2135 GTGATGAGTGAAGATGAT 2154
 DB 87931 GTGATGAGTGAAGATGAT 87950

RESULT 8
 US-09-109-063-26
 Sequence 26, Application US/09109063
 Patent No. 6013498
 GENERAL INFORMATION:
 APPLICANT: YOKOYAMA, KEIICHI
 APPLICANT: NAKAMURA, NAMI
 APPLICANT: MIWA, TETSUYA
 APPLICANT: SEGURO, KATSUYA
 TITLE OF INVENTION: PROCESS FOR PRODUCING MICROBIAL TRANSGLUTAMINASE
 FILE REFERENCE: 0010-0937-0
 CURRENT APPLICATION NUMBER: US/09/109, 063
 EARLIER FILING DATE: 1998-07-02
 EARLIER APPLICATION NUMBER: JP 180010/1997
 NUMBER OF SEQ ID NOS: 62
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 26
 LENGTH: 45
 TYPE: DNA
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence:SYNTHETIC DNA
 US-09-109-063-26

Query Match 0.68; Score 19; DB 3; Length 45;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 490 GAGACGCTCATGATGAT 508
 DB 10 GAGACGCTCATGATGAT 28

RESULT 9
 US-09-109-063-27/c
 Sequence 27, Application US/09109063
 Patent No. 6013498
 GENERAL INFORMATION:
 APPLICANT: YOKOYAMA, KEIICHI
 APPLICANT: NAKAMURA, NAMI
 APPLICANT: MIWA, TETSUYA
 APPLICANT: SEGURO, KATSUYA
 TITLE OF INVENTION: PROCESS FOR PRODUCING MICROBIAL TRANSGLUTAMINASE
 FILE REFERENCE: 0010-0937-0

;; CURRENT APPLICATION NUMBER: US/09/109,063
;; CURRENT FILING DATE: 1998-07-02
;; EARLIER APPLICATION NUMBER: JP 180010/1997
;; EARLIER FILING DATE: 1997-07-04
;; NUMBER OF SEQ ID NOS: 62
;; SOFTWARE: Patentin Ver. 2.0
;; LENGTH: 50
;; TYPE: DNA
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC DNA
US-09-109-063-27
Query Match 0.6%; Score 19; DB 3; Length 50;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 490 GAGAGCGCTCATGATGAAT 508
DB 45 GAGAGCGCTCATGATGAAT 27
RESULT 10
US-09-109-063-2
;; Sequence 2, Application US/09109063
;; Patent No. 6013498
;; GENERAL INFORMATION:
;; APPLICANT: YOKOTAMA, KEIICHI
;; APPLICANT: NAKAMURA, NAHI
;; APPLICANT: MIWA, TETSUYA
;; APPLICANT: SEGURO, KATSUYA
;; TITLE OF INVENTION: PROCESS FOR PRODUCING MICROBIAL TRANSGLUTAMINASE
;; FILE REFERENCE: 0010-0937-0
;; CURRENT APPLICATION NUMBER: US/09/109,063
;; CURRENT FILING DATE: 1998-07-02
;; EARLIER APPLICATION NUMBER: JP 180010/1997
;; EARLIER FILING DATE: 1997-07-04
;; NUMBER OF SEQ ID NOS: 62
;; SOFTWARE: Patentin Ver. 2.0
;; SEQ ID NO 2
;; LENGTH: 993
;; TYPE: DNA
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC DNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)..(993)
;; OTHER INFORMATION: IDENTIFICATION METHOD: S
US-09-109-063-2
Query Match 0.6%; Score 19; DB 3; Length 993;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 490 GAGAGCGCTCATGATGAAT 508
DB 412 GAGAGCGCTCATGATGAAT 430
RESULT 11
US-08-416-478A-4/C
;; Sequence 4, Application US/08416478A
;; Patent No. 5773578
;; GENERAL INFORMATION:
;; APPLICANT: Hercend, Thierry
;; APPLICANT: Triebel, Frederic
;; TITLE OF INVENTION: New proteins produced by human lymphocytes, DNA sequences encoding these proteins and their pharmaceutical and biological uses
;; NUMBER OF SEQUENCES: 10
;; CORRESPONDENCE ADDRESS:

ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
STREET: 419 Seventh Street, N.W., Suite 400
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/416,478A
FILING DATE: 04-APR-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/416,478
FILING DATE: 04-APR-1995
CLASSIFICATION: 435
US-08-416-478A-4
Query Match 0.6%; Score 19; DB 1; Length 999;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 880 TGGGGCATGAGAGATRACT 898
DB 140 TGGGGCATGAGAGATRACT 122
RESULT 12
US-08-474-988B-4/C
;; Sequence 4, Application US/08474988B
;; Patent No. 5874250
;; GENERAL INFORMATION:
;; APPLICANT: Hercend, Thierry
;; APPLICANT: Triebel, Frederic
;; TITLE OF INVENTION: NEW PROTEINS PRODUCED BY HUMAN LYMPHOCYTES, DNA SEQUENCES ENCODING THESE PROTEINS AND THEIR PHARMACEUTICAL AND BIOLOGICAL USES
;; NUMBER OF SEQUENCES: 11
;; CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.
STREET: 419 Seventh Street, N.W., Suite 400
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,988B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/416,478
FILING DATE: 04-APR-1995
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/854,644
FILING DATE: 08-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: YUN, Allen C.
REGISTRATION NUMBER: 37,971
REFERENCE/DOCKET NUMBER: HERCEND-1B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 999 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-474-988B-4

Query Match 0.6%; Score 19; DB 2; Length 19;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 880 TGGGCATGAGATAACT 898
|||||
DB 140 TGGGCATGAGATAACT 122

RESULT 13
US-08-394-442B-4/C
Sequence 4, Application US/08394442B
Patent No. 5976877
GENERAL INFORMATION:
APPLICANT: Hercend, Thierry
TITLE OF INVENTION: New Proteins Produced by Human
TITLE OF INVENTION: Lymphocytes, DNA Sequences Produced by These Proteins And
TITLE OF INVENTION: Their Pharmaceutical And Biological Uses
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEWMARK, P.L.L.C.
STREET: 419 Seventh Street, N.W., Suite 400
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/394,442B
FILING DATE: 24-FEB-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: HERCEND-2A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 999 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-394-442B-4

Query Match 0.6%; Score 19; DB 2; Length 19;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 880 TGGGCATGAGATAACT 898
|||||
DB 140 TGGGCATGAGATAACT 122

RESULT 14
PCT-US91-04274A-4
Sequence 4, Application PC/TUS9104274A
GENERAL INFORMATION:
APPLICANT: Williams, Douglas E.
APPLICANT: Lyman, Stewart
TITLE OF INVENTION: Mast Cell Growth Factor
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/04274A
FILING DATE: 19910614
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Oster, Jeffrey B.
REGISTRATION NUMBER: 32585
REFERENCE/DOCKET NUMBER: 0521D
TELECOMMUNICATION INFORMATION:
TELEPHONE: 2065870430
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1050 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: N
ORIGINAL SOURCE:
ORGANISM: HUMAN MAST CELL GROWTH FACTOR
IMMEDIATE SOURCE:
CLONE: HMGF-2.4
PCT-US91-04274A-4

Query Match 0.6%; Score 19; DB 5; Length 1050;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1214 AGAAGAAAAACATCTGA 1232
|||||
DB 574 AGAAGAAAAACATCTGA 592

RESULT 15
PCT-US91-04274A-3
Sequence 3, Application PC/TUS9104274A
GENERAL INFORMATION:
APPLICANT: Williams, Douglas E.
APPLICANT: Lyman, Stewart
TITLE OF INVENTION: Mast Cell Growth Factor
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101

```

:
: COMPUTER READABLE FORM:
:
: MEDIUM TYPE: floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.24
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US91/04274A
: FILING DATE: 19910614
:
: CLASSIFICATION: 435
:
: ATTORNEY/AGENT INFORMATION:
: NAME: Oster, Jeffrey B.
: REGISTRATION NUMBER: 32585
: REFERENCE/DOCKET NUMBER: 0521D
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 2065870430
:
: INFORMATION FOR SEQ ID NO: 3:
:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1069 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHEetical: N
: ANTI-SENSE: N
:
: ORIGINAL SOURCE:
: ORGANISM: HUMAN MAST CELL GROWTH FACTOR
: IMMEDIATE SOURCE:
: CLONE: HMGF-2D
:
: PCT-US91-04274A-3
:
: Query Match 0.6%; Score 19; DB 5; Length 1159;
: Best Local Similarity 100.0%; Pred. No. 26;
: Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
:
: QY 1214 AGAAGAAAAAGCATCTGA 1232
: ||||||||||||||||
: DB 575 AGAAGAAAAAGCATCTGA 593

```

Search completed: May 16, 2003, 20:08:09
 Job time : 1831 secs

PN MO200164707-A1.
 XX
 PD 07-SEP-2001.
 XX
 PF 28-FEB-2001; 2001WO-US06960.
 XX
 PR 29-FEB-2000; 2000US-0515363.
 XX
 PA (UYCO) UNIV COLUMBIA NEW YORK.
 XX
 PI Fisher PB, Kang D, GopalKrishnan RV;
 XX
 DR MPI: 2001-565494/63.
 DR P-PSDB: AAEI0155.
 XX
 PT Nucleic acid sequences encoding a Melanoma Differentiation Associated
 PT Gene useful for cancer cell growth suppression, apoptosis and
 PT anti-viral activity -
 XX
 PS Claim 1: Page 16-18; 152pp; English.
 XX
 CC The present invention relates to an isolated nucleic acid encoding a
 CC melanoma differentiation associated gene-5 (Mda-5) polypeptide. Mda-5
 CC contains a caspase recruitment domain (CARD) and a RNA helicase motif.
 CC Mda-5 is a novel interferon (IFN) inducible gene with structural
 CC similarities to RNA helicases and CARD motif containing proteins. Mda-5
 CC is induced during terminal differentiation in human melanoma cells
 CC treated with the combination of recombinant fibroblast IFN and the
 CC antileukemic compound mezerein (MEZ). Mda-5 is useful for identifying
 CC compounds that may induce its expression. Mda-5 is useful for treating
 CC cancer e.g., melanoma, neuroblastoma, astrocytoma, glioblastoma
 CC multiforme, cervical cancer, breast cancer, colon cancer, prostate
 CC cancer, osteosarcoma, chondrosarcoma and especially where the cancer is
 CC a cancer of the central nervous system and apoptosis. The Mda-5 promoter
 CC exhibits melanocyte tissue specificity and minimizes systemic toxicity.
 CC The present sequence is human Mda-5 cDNA.
 XX
 SO Sequence 3365 BP; 1138 A; 644 C; 753 G; 830 T; 0 other:
 Query Match 100.0%; Score 3365; DB 22; Length 3365;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 421 AGACCGGAGCCCTCTGGCCGCCCGCTACATGAAACCTGAGCTCAGGACTTCCCTCT 480
 QY 481 CCATGTTTGAAGAGCTCATGATGATATATCCAACTGTGAACCTCTTCAAGCCACT 540
 Db 481 CCATGTTTGAAGAGCTCATGATGATATATCCAACTGTGAACCTCTTCAAGCCACT 540
 QY 541 CTGTGTGACAGCTTCTAGTATAGAGAGCTTTGATTAAGTGCAATGAGAGAGAACTGTTG 600
 Db 541 CTGTGTGACAGCTTCTAGTATAGAGAGCTTTGATTAAGTGCAATGAGAGAGAACTGTTG 600
 QY 601 ACAATTGAAGAGAGAAACCGGATTCGTGTCAGAAAACCAATGAAATGAATCAAGTGA 660
 Db 601 ACAATTGAAGAGAGAAACCGGATTCGTGTCAGAAAACCAATGAAATGAATCAAGTGA 660
 QY 661 AGAGAGCTACTAAAAAGATTGTGAGAAAGAAACCTGTTCTGCAATTTCTGAATGT 720
 Db 661 AGAGAGCTACTAAAAAGATTGTGAGAAAGAAACCTGTTCTGCAATTTCTGAATGT 720
 QY 721 CTTCGTCAAGAGAGAAACCAATGATTTGTCAGAGGTTTACAGGCTGTGATTTGTCAGAA 780
 Db 721 CTTCGTCAAGAGAGAAACCAATGATTTGTCAGAGGTTTACAGGCTGTGATTTGTCAGAA 780
 QY 781 ACCAATGACAGATTGAGATTTATACAGAGTTGATGCTCAAGTGAAGAGCACTT 840
 Db 781 ACCAATGACAGATTGAGATTTATACAGAGTTGATGCTCAAGTGAAGAGCACTT 840
 QY 841 CTTTCAACCCAGTTTCAGCAATCTGAGAGAGAGTCTGGGCAATGAGAAATTAATCA 900
 Db 841 CTTTCAACCCAGTTTCAGCAATCTGAGAGAGAGTCTGGGCAATGAGAAATTAATCA 900
 QY 901 TCAGATATCTTTTGGCAGATTTCTTCTAGTTAGATTAAGATGACACAAAGTTTGGCAAA 960
 Db 901 TCAGATATCTTTTGGCAGATTTCTTCTAGTTAGATTAAGATGACACAAAGTTTGGCAAA 960
 QY 961 GGAATGTCAGCTGCTTGTGATGAAGTTTGAACATTAACAGCAATGAGAGAGTATCA 1020
 Db 961 GGAATGTCAGCTGCTTGTGATGAAGTTTGAACATTAACAGCAATGAGAGAGTATCA 1020
 QY 1021 GGCACCATGAGAGATGATTCAGATGAAGAAATGTGGCAGCAAGCATCCCGGAGCCA 1080
 Db 1021 GGCACCATGAGAGATGATTCAGATGAAGAAATGTGGCAGCAAGCATCCCGGAGCCA 1080
 QY 1081 GAACTCCAGCTGAGGCTTACCAATGAAATGAGTTGCGCAGCAAGCTTGGAGGAGAA 1140
 Db 1081 GAACTCCAGCTGAGGCTTACCAATGAAATGAGTTGCGCAGCAAGCTTGGAGGAGAA 1140
 QY 1141 ATCATCATCTGCTCCCTACAGAGAGTGAAGAAACAGAGTGGCTTTACATTTCCCAAG 1200
 Db 1141 ATCATCATCTGCTCCCTACAGAGAGTGAAGAAACAGAGTGGCTTTACATTTCCCAAG 1200
 QY 1201 GATCAGCTTAGACAGAGAAAAAGCATCTGAGCTTGAAAAATATATGTTCTTCTCAAT 1260
 Db 1201 GATCAGCTTAGACAGAGAAAAAGCATCTGAGCTTGAAAAATATATGTTCTTCTCAAT 1260
 QY 1261 AAGGTACGTGATGTTGAACAGCTCTTCGCAAGAGTTCCAAACCTTTTGAAGAAATGG 1320
 Db 1261 AAGGTACGTGATGTTGAACAGCTCTTCGCAAGAGTTCCAAACCTTTTGAAGAAATGG 1320
 QY 1321 TATCGTTATATGATTAAGTGTGATACCAACAGAAATATCATTTCCAGAAATGTC 1380
 Db 1321 TATCGTTATATGATTAAGTGTGATACCAACAGAAATATCATTTCCAGAAATGTC 1380
 QY 1381 AAGTCTGTGATATATATATAGTACAGCTCAAAATCCTTGAAGAACTCCCTTAACTTG 1440
 Db 1381 AAGTCTGTGATATATATATAGTACAGCTCAAAATCCTTGAAGAACTCCCTTAACTTG 1440
 QY 1441 GAAATGAGAGAGATGCTGCTTCAATTTGTCAGACTTTTCCCTCATTTATATGATGAA 1500
 Db 1441 GAAATGAGAGAGATGCTGCTTCAATTTGTCAGACTTTTCCCTCATTTATATGATGAA 1500
 QY 1501 TGTATACACCAACAAAGAGAGTGTATATATACATCATGAGGCAATATTTATGAG 1560
 Db 1501 TGTATACACCAACAAAGAGAGTGTATATATACATCATGAGGCAATATTTATGAG 1560

[illegible]

```

FT      /tag- a
PI      /product- "Human RH16"
PN      WO200185955-A1.
PD      15-NOV-2001.
XX      11-MAY-2001: 2001WO-FR01441.
XX      11-MAY-2000: 2000FR-0006030.
XX      (ISTA-) INSTAC.
XX      (INSP) INST PASTEUR LILLE.
XX      Bahr G, Cocude C, Capron A;
XX      WPI: 2002-082898/11.
XX      P-PSDB: AAM47798.
PT      New polypeptide, useful for treating and diagnosing cancer or
PT      inflammation, and drug screening, comprises a human polynucleotide
PT      homologous to RNA helicase
XX      PS
XX      Claim 7: Page 85-89; 114pp: French.
XX      CC
XX      The present sequence is the coding sequence for human RH16. RH16 is a
XX      116kDa protein and has homology to RNA helicases (DEXH box). RH16 and
XX      its coding sequence are useful for treating cancer; acute or chronic
XX      infections (especially by HIV or hepatitis B or C); inherited genetic
XX      diseases (autoimmune diseases (particularly rheumatism, arthritis,
XX      arteriosclerosis, osteoporosis and diabetes, but many others listed),
XX      and to prevent graft rejection. RH16 and its coding sequence are also useful
XX      for inducing, or increasing, the immune response to a vaccine.
XX      CC
XX      Sequence 3372 BP; 1157 A; 637 C; 748 G; 830 T; 0 other:
SQ
Query Match      90.4%; Score 3042; DB 24; Length 3372;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 3342; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

```

DB      483 CTCATGATGAATATCTCCAACTGCTGAACCTCTTCAACCCACCTGCTGACAGCTTG 542
OY      TAGTAGAGACGCTCTTGATAGTCATAGAGAGAGAACTCTGCATATTCAGACAGAA 616
DB      543 TAGTAGAGACGCTCTTGATAGTCATAGAGAGAGAACTCTGCAATTCAGACAGAA 602
OY      ACCGATTTGCTGTCAGAGAAACATGGAATGAAATGAGTGTAGAGAGCTACTAAAAA 676
DB      603 ACCGATTTGCTGTCAGAGAAACATGGAATGAAATGAGTGTAGAGAGCTACTAAAAA 662
OY      GGATTTGTCAGAGAAACATGGAATGAAATGAGTGTAGAGAGCTACTAAAAA 736
DB      663 GGATTTGTCAGAGAAACATGGAATGAAATGAGTGTAGAGAGCTACTAAAAA 722
OY      ACAATGAACTTGTCCAGAGAGTTCAGAGAGCTCTGATTTGCTAGAGAAAGCAATGAGATG 796
DB      723 ACAATGAACTTGTCCAGAGAGTTCAGAGAGCTCTGATTTGCTAGAGAAAGCAATGAGATG 782
OY      AGAATTTATCAGAGAGTTCAGAGAGCTCTGATTTGCTAGAGAAAGCAATGAGATG 856
DB      783 AGAATTTATCAGAGAGTTCAGAGAGCTCTGATTTGCTAGAGAAAGCAATGAGATG 842
OY      AGCCAAATCTGAGAGAGAGTTCAGAGAGCTCTGATTTGCTAGAGAAAGCAATGAGATG 916
DB      843 AGCCAAATCTGAGAGAGAGTTCAGAGAGCTCTGATTTGCTAGAGAAAGCAATGAGATG 902
OY      CAGATTTCTGCTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 976
DB      903 CAGATTTCTGCTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTTG 962
OY      TAGATGAAAGCTTGTGATTAACAGCAACATGAGGAGTATTCAGAGAGCTACTGAAAGTG 1036
DB      963 TAGATGAAAGCTTGTGATTAACAGCAACATGAGGAGTATTCAGAGAGCTACTGAAAGTG 1022
OY      ATTGATGAAAGAGTATGATTAACAGCAACATGAGGAGTATTCAGAGAGCTACTGAAAGTG 1096
DB      1033 ATTGATGAAAGAGTATGATTAACAGCAACATGAGGAGTATTCAGAGAGCTACTGAAAGTG 1082
OY      CTACCAATGGAAGTTCAGAGAGCTCTGATTTGATTTGATTTGATTTGATTTGATTTGAT 1156
DB      1083 CTACCAATGGAAGTTCAGAGAGCTCTGATTTGATTTGATTTGATTTGATTTGATTTGAT 1142
OY      CTACAGGAGTGTGAGAAACAGAGAGTGTATTTGATTTGATTTGATTTGATTTGATTTG 1216
DB      1143 CTACAGGAGTGTGAGAAACAGAGAGTGTATTTGATTTGATTTGATTTGATTTGATTTG 1202
OY      AGAAAAAGCATGAGAGCTGTGAGAAAGTATTTGATTTGATTTGATTTGATTTGATTTG 1276
DB      1203 AGAAAAAGCATGAGAGCTGTGAGAAAGTATTTGATTTGATTTGATTTGATTTGATTTG 1262
OY      AACACCTCTTCCGCAAGAGCTTCCAGACATTTTGAAGAAATGATGATTTGATTTGAT 1336
DB      1263 AACACCTCTTCCGCAAGAGCTTCCAGACATTTTGAAGAAATGATGATTTGATTTGAT 1322
OY      TAAGTGTGATTAACCACTGAAATATATTTCCAGAAAGTGTCAAGTCTGTGATTTA 1396
DB      1337 TAAGTGTGATTAACCACTGAAATATATTTCCAGAAAGTGTCAAGTCTGTGATTTA 1382
OY      TAAGTGTGATTAACCACTGAAATATATTTCCAGAAAGTGTCAAGTCTGTGATTTA 1362
DB      1363 TAAGTGTGATTAACCACTGAAATATATTTCCAGAAAGTGTCAAGTCTGTGATTTA 1356
OY      TTATGATGAGAGCTCAAAATCTGAAAGCTCTTGAATTTGAATTTGAATTTGAATTTGA 1456
DB      1457 TTATGATGAGAGCTCAAAATCTGAAAGCTCTTGAATTTGAATTTGAATTTGAATTTGA 1442
OY      CTGATGATTAATTTGATTAATTTGATTAATTTGATTAATTTGATTAATTTGATTAATTT 1516
DB      1443 CTGATGATTAATTTGATTAATTTGATTAATTTGATTAATTTGATTAATTTGATTAATTT 1502
OY      AAGAAGAGTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1576
DB      1503 AAGAAGAGTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1562
OY      GAGTCAAGAAAGAAACCAAGAGTATGCTCTTCAATGATGAGATTAATTAATTAATTAAT 1636
DB      1577 GAGTCAAGAAAGAAACCAAGAGTATGCTCTTCAATGATGAGATTAATTAATTAATTAATTA

```

Db 1563 GACTCAAGAAAAAACAACCCAGTGTATCCCTTCCTCAGATACCTGGAATTAACAGCTT 1622
 QY 1637 CACCTGCTGTGGAGGGGCGCAGAGCAAGCCAAAGCTCAATACACACATTAACATAT 1696
 Db 1623 CACCTGCTGTGGAGGGGCGCAGAGCAAGCCAAAGCTCAATACACACATTAACATAT 1682
 QY 1697 GTGCCAATCTGTACATTTACTATTAAACCTGTAAACAAACCTGTAACTGAAA 1756
 Db 1683 GTGCCAATCTGTACATTTACTATTAAACCTGTAAACAAACCTGTAACTGAAA 1742
 QY 1757 ACCAAATACAGAGCCATGCAAGAGTTGGCATTGCAGATGCCAACAGTAAGATCCAT 1816
 Db 1743 ACCAAATACAGAGCCATGCAAGAGTTGGCATTGCAGATGCCAACAGTAAGATCCAT 1802
 QY 1817 TTAAGAGAACTCTGAGAAATTAATGACAGAGATCAAACTTATGTCAATGATCCAA 1876
 Db 1803 TTAAGAGAACTCTGAGAAATTAATGACAGAGATCAAACTTATGTCAATGATCCAA 1862
 QY 1877 TGTGAGATTTTGAATCAACCCCTATGAAACAATGGGGCCATTCAATGGAAATAAAAGCTG 1936
 Db 1863 TGTGAGATTTTGAATCAACCCCTATGAAACAATGGGGCCATTCAATGGAAATAAAAGCTG 1922
 QY 1937 CAAAAAAGAGAAATGCAAGAAAGCTGTTGTGCAGAACATTTGAGGAATACAAATGAG 1996
 Db 1923 CAAAAAAGAGAAATGCAAGAAAGCTGTTGTGCAGAACATTTGAGGAATACAAATGAG 1982
 QY 1997 CCTCAAAATTAATGACAAATTCGAAATGATGCGATCTCATCTCCAACTTCT 2056
 Db 1983 CCTCAAAATTAATGACAAATTCGAAATGATGCGATCTCATCTCCAACTTCT 2042
 QY 2057 ATAAATGAGAGAAAGATTAAGAAAGTTGCGATGATGAAATGATGATGATGATGATGATG 2116
 Db 2043 ATAAATGAGAGAAAGATTAAGAAAGTTGCGATGATGAAATGATGATGATGATGATGATG 2102
 QY 2117 ATGATGAGATTTGTGATGCTGATGAAGATGAGATGATTTAAAGAAACCTCTGAAATG 2176
 Db 2103 ATGATGAGATTTGTGATGCTGATGAAGATGAGATGATTTAAAGAAACCTCTGAAATG 2162
 QY 2177 ATGAAACAGATGATTTGATGATGATTTTGAACAAATGAAATTTGAAAGGC 2236
 Db 2163 ATGAAACAGATGATTTGATGATGATTTTGAACAAATGAAATTTGAAAGGC 2222
 QY 2237 TGGCTGAAAAACCCAGATATGAAAAATGAAAGCTGACCAATTAAGAAAATACATATG 2296
 Db 2223 TGGCTGAAAAACCCAGATATGAAAAATGAAAGCTGACCAATTAAGAAAATACATATG 2282
 QY 2297 AGCATATATCTAGAGCTAGAGAAATGACAGAGAGAAATATCTTTACAAATACACAGACA 2356
 Db 2283 AGCATATATCTAGAGCTAGAGAAATGACAGAGAGAAATATCTTTACAAATACACAGACA 2342
 QY 2357 GTGCATATGCGGTTTCCAGTGGATTAAGAAATGAAATTTGCTGATGAGAGTCA 2416
 Db 2343 GTGCATATGCGGTTTCCAGTGGATTAAGAAATGAAATTTGCTGATGAGAGTCA 2402
 QY 2417 AAGCCCAACATCTGATGAGCTGAGACAGAGAGTGTCAAAACCATTAACAGAAATG 2476
 Db 2403 AAGCCCAACATCTGATGAGCTGAGACAGAGAGTGTCAAAACCATTAACAGAAATG 2462
 QY 2477 AACAAAAAGAGCTAGTAAATTTCCAGCTGAGAAAAATCAATCTGCTTATCGTACCA 2536
 Db 2463 AACAAAAAGAGCTAGTAAATTTCCAGCTGAGAAAAATCAATCTGCTTATCGTACCA 2522
 QY 2537 CAGTGCAGAGAGAGTCTGATTAAGAAATGTAACTGTTATCCGATATGCTCTG 2596
 Db 2523 CAGTGCAGAGAGAGTCTGATTAAGAAATGTAACTGTTATCCGATATGCTCTG 2582
 QY 2597 TCACCAATGAATAGCCATGCTCAGAGCCGCTGCTGAGCCAGAGCTGATAGAGACCT 2656
 Db 2583 TCACCAATGAATAGCCATGCTCAGAGCCGCTGCTGAGCCAGAGCTGATAGAGACCT 2642
 QY 2657 AGCTCTGCTGCTCAGAGTGTGTTGAGAGATTAATGAGATGAGACATGATGATTC 2716
 Db 2643 AGCTCTGCTGCTCAGAGTGTGTTGAGAGATTAATGAGATGAGACATGATGATTC 2702

QY 2717 GAGAGAAATGATGATTAAGCTATACATTTGTTCAAAATATGAAACAGAGAAATATG 2776
 Db 2703 GAGAGAAATGATGATTAAGCTATACATTTGTTCAAAATATGAAACAGAGAAATATG 2762
 QY 2777 CTCATTAAGATTTTGAATTTAGAGATGCAAGATTAATGAAAGAAATGAAATGAA 2836
 Db 2763 CTCATTAAGATTTTGAATTTAGAGATGCAAGATTAATGAAAGAAATGAAATGAA 2822
 QY 2837 GAAATTTTGGCAAGCATTTACAGAAATTAACCATTAATTAATCTTCTGTTGCAAAACT 2896
 Db 2823 GAAATTTTGGCAAGCATTTACAGAAATTAACCATTAATTAATCTTCTGTTGCAAAACT 2882
 QY 2897 GCAGTGTCTAGCCCTGTTCTGGGAAATTCATGATTTGAGAAAAATGATCAAGTCA 2956
 Db 2883 GCAGTGTCTAGCCCTGTTCTGGGAAATTCATGATTTGAGAAAAATGATCAAGTCA 2942
 QY 2957 ATATGACCCGAGAAATTCAGAGAACTTTTACATTTGTAAGAAAAACAAAGCACTGCAGAA 3016
 Db 2943 ATATGACCCGAGAAATTCAGAGAACTTTTACATTTGTAAGAAAAACAAAGCACTGCAGAA 3002
 QY 3017 AGTGTGCGGACTATCAATTAATGTTGAATTCATCTGCAATTTGCGCAGCTTGCGGAA 3076
 Db 3003 AGTGTGCGGACTATCAATTAATGTTGAATTCATCTGCAATTTGCGCAGCTTGCGGAA 3062
 QY 3077 CAATGATGTTGCAAGAGGCTTAGATTTGCCCTGCTCAAAATTAAGAAATTTTGTAGTGG 3136
 Db 3063 CAATGATGTTGCAAGAGGCTTAGATTTGCCCTGCTCAAAATTAAGAAATTTTGTAGTGG 3122
 QY 3137 TTTTCAAAATTAATTCACAAAGAAACAAATACAAAAGTGGGTAGAAATTAACAT 3196
 Db 3123 TTTTCAAAATTAATTCACAAAGAAACAAATACAAAAGTGGGTAGAAATTAACAT 3182
 QY 3197 TTCCCAATCTTGTACTATTCAGAAATGCTGTTATTTAGATGAGAGATTAAGCACTGATG 3256
 Db 3183 TTCCCAATCTTGTACTATTCAGAAATGCTGTTATTTAGATGAGAGATTAAGCACTGATG 3242
 QY 3257 AAGATTTCTTTAAATTAATTAATCAATTAATTAATTAATTAATTAATTAATTAAT 3316
 Db 3243 AAGATTTCTTTAAATTAATTAATCAATTAATTAATTAATTAATTAATTAATTAAT 3302
 QY 3317 TATGCTACAGAACTGACATTAAGAAATCAATTAATTAATTAATTAATTAATTAAT 3364
 Db 3303 TATGCTACAGAACTGACATTAAGAAATCAATTAATTAATTAATTAATTAATTAAT 3350

RESULT 3
 AAD11170
 ID AAD11170 standard; DNA: 3131 BP.
 AC AAD11170;
 XX 29-NOV-2001 (first entry)
 DT XX
 DE XX
 XX Human melanoma differentiation associated-5 protein-related (M5).
 DE Human melanoma differentiation associated gene; M5-5; Interferon; IFN;
 KW RNA helicase motif; caspase recruitment domain; CARD; therapy; melanoma;
 KW neuroblastoma; astrocytoma; glioblastoma; multiforme; cancer; cervical;
 KW breast; colon; prostate; osteosarcoma; chondrosarcoma; systemic toxicity;
 KW central nervous system; cytosolic; apoptosis; ds.
 XX OS
 XX Homo sapiens.
 XX PN
 XX MO200164707-A1.
 PD 07-SEP-2001.
 XX 28-FEB-2001; 2001WO-US06960.
 PF 29-FEB-2000; 2000US-0515363.
 XX PR
 XX (UNCO) UNIV COLUMBIA NEW YORK.

XX Fisher PB, Kang D, GopalKrishnan RV;
 XX WPI; 2001-565494/63.
 XX
 PT Nucleic acid sequences encoding a Melanoma Differentiation Associated
 PT Gene useful for cancer cell growth suppression, apoptosis and
 PT anti-viral activity
 XX
 XX Disclosure: Page 134-148; 152pp: English.
 CC
 CC The present invention relates to an isolated nucleic acid encoding a
 CC melanoma differentiation associated gene-5 (Mda-5) polypeptide. Mda-5
 CC contains a caspase recruitment domain (CARD) and a RNA helicase motif.
 CC Mda-5 is a novel interferon (IFN) inducible gene with structural
 CC similarities to RNA helicases and CARD motif containing proteins. Mda-5
 CC is induced during terminal differentiation in human melanoma cells
 CC treated with the combination of recombinant fibroblast IFN and the
 CC anti-leukemic compound mezerein (MEZ). Mda-5 is useful for identifying
 CC compounds that may induce its expression. Mda-5 is useful for treating
 CC cancer e.g., melanoma, neuroblastoma, astrocytoma, glioblastoma
 CC multiforme, cervical cancer, breast cancer, colon cancer, prostate
 CC cancer, osteosarcoma, chondrosarcoma and especially where the cancer is
 CC a cancer of the central nervous system and apoptosis. The Mda-5 promoter
 CC exhibits melanocyte tissue specificity and minimizes systemic toxicity.
 CC The present sequence is human Mda-5 protein-related DNA.
 CC Note: The present sequence is designated SEQ ID NO:2 in the sequence
 CC listing, but does not correspond to the sequence designated SEQ ID NO:2
 CC in the main body of the specification (AAE10155). The present sequence is
 CC not further referred to in the specification, and has been represented
 CC in a protein format in the sequence listing.
 XX
 XX Sequence 3131 BP; 1075 A; 583 C; 699 G; 774 T; 0 other;
 SO
 Query Match 80.5%; Score 2709; DB 22; Length 3131;
 Best Local Similarity 99.8%; Pred No. 0;
 Matches 3129; Conservative 0; Mismatches 0; Indels 5; Gaps 2;

QY 709 TTTTGAATGTTCTTCGTCACAAACAGAAACATGAACTGTGTCACAGTAAACAGCTCT 768
 DB 541 TTTTGAATGTTCTTCGTCACAAACAGAAACATGAACTGTGTCACAGTAAACAGCTCT 600
 QY 769 GATTGCTAGAAAGCAATGACAGAGATTGAGATTTATGACAAAGTTGATGCTCTAGCTG 828
 DB 601 GATTGCTAGAAAGCAATGACAGAGATTGAGATTTATGACAAAGTTGATGCTCTAGCTG 660
 QY 829 GAAAGCAAGCTCTTCACACACAGATTGAGCAAAATGAGAGAGAGCTGGGCTATG 888
 DB 661 GAAAGCAAGCTCTTCACACACAGATTGAGCAAAATGAGAGAGAGCTGGGCTATG 720
 QY 889 GAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 948
 DB 721 GAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 780
 QY 949 AGTTTGSCAGAAAGAGTGCAGCTGCTTACATGAAAGCTTTGACATACAGCAATATG 1008
 DB 781 AGTTTGSCAGAAAGAGTGCAGCTGCTTACATGAAAGCTTTGACATACAGCAATATG 840
 QY 1009 GGCAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1068
 DB 841 GGCAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
 QY 1069 TCCCGGAGCCAGAACTGACCTGAGGCTTACCAAAATGAGAGTGGCCACACCAACCTTG 1128
 DB 901 TCCCGGAGCCAGAACTGACCTGAGGCTTACCAAAATGAGAGTGGCCACACCAACCTTG 960
 QY 1129 GAAAGGAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1188
 DB 961 GAAAGGAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1020
 QY 1189 TACATTTGCCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1248
 DB 1021 TACATTTGCCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1080
 QY 1249 GTTCTTGCAATTAAGTACTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1308
 DB 1081 GTTCTTGCAATTAAGTACTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1140
 QY 1309 TTGAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1368
 DB 1141 TTGAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1200
 QY 1369 CCAGAAAGTTCAGAACTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1428
 DB 1201 CCAGAAAGTTCAGAACTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1257
 QY 1429 CTCTTAAAGTTGAAATGAGAAAGATGCTGTTCAATTTGTGACACTTTTCCATTT 1488
 DB 1258 CTCTTAAAGTTGAAATGAGAAAGATGCTGTTCAATTTGTGACACTTTTCCATTT 1317
 QY 1489 ATCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1548
 DB 1318 ATCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1377
 QY 1549 TATTGATGAGAAAGTTCAGAAAGTTCAGAAAGTTCAGAAAGTTCAGAAAGTTCAGAAAGTTC 1608
 DB 1378 TATTGATGAGAAAGTTCAGAAAGTTCAGAAAGTTCAGAAAGTTCAGAAAGTTCAGAAAGTTC 1437
 QY 1609 CTCTCTGATGATGAGAAAGTTCAGAAAGTTCAGAAAGTTCAGAAAGTTCAGAAAGTTCAGAA 1668
 DB 1438 CTCTCTGATGATGAGAAAGTTCAGAAAGTTCAGAAAGTTCAGAAAGTTCAGAAAGTTCAGAA 1497
 QY 1669 AAAGTGAAGAAACATTTTAAACATATGCTGATGATGATGATGATGATGATGATGATGATG 1728
 DB 1498 AAAGTGAAGAAACATTTTAAACATATGCTGATGATGATGATGATGATGATGATGATGATG 1557
 QY 1729 GTTAAAGAAAGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1788
 DB 1558 GTTAAAGAAAGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1617

OY	1789	ATCCACATGCAACCCAGAGAGATCCATTTAAAGAAACCTTTAGAA/ATGCAAG	1848
Db	1618	ATTCACATGCAACCCAGAGAGATTCATTTAAAGAAACCTTTAGAA/ATGCAAG	1677
OY	1849	ATTCAAACCTATTGTCMAATAGAGTCACATGTCAGATTTGCAATTCAC/ATGCAAG	1908
Db	1678	ATTCAAACCTATTGTCMAATAGAGTCACATGTCAGATTTGCAATTCAC/ATGCAAG	1737
OY	1909	TGGGCCATTCMAATGCAAAAAAAAAAGCTGCAAAAAAAAAAGAAATCGCAAAAGAGTGTGT	1968
Db	1738	TGGGCCATTCMAATGCAAAAAAAAAAGCTGCAAAAAAAAAAGAAATCGCAAAAGAGTGTGT	1797
OY	1969	GCACAACTTTGAGAGAGTCACATGAGGCCCTTCACAAATTAAGACACAT/AGAAATGA	2028
Db	1798	GCACAACTTTGAGAGAGTCACATGAGGCCCTTCACAAATTAAGACACAT/AGAAATGA	1857
OY	2029	GAT/-GGGTATACGATCTTGAACTTTCTATATGAGAGAGAAATAAAGTTTGAG	2086
Db	1858	GATCGGGTATACGATCTTGAACTTTCTATATGAGAGAGAAATAAAGTTTGAG	1917
OY	2087	TCATAGAAATGATGATGATGAGGGGTGGTATGATGATGATTTGTATGAG/ATGAGATG	2146
Db	1918	TCATAGAAAGATGATGATGATGAGGGGTGGTATGATGATGATTTGTATGAG/ATGAGATG	1977
OY	2147	AGGATGATTTAAAGAAACCTTTGAAACTGGATGAAACACATAGATTCT/AGATTTAT	2206
Db	1978	AGGATGATTTAAAGAAACCTTTGAAACTGGATGAAACACATAGATTCT/AGATTTAT	2037
OY	2207	TTTTTGAAGAAACAATAAATTTGTAAGAAAGCTGGCTCAAAACCCAGATATCAAAATGAA	2266
Db	2038	TTTTTGAAGAAACAATAAATTTGTAAGAAAGCTGGCTCAAAACCCAGATATCAAAATGAA	2097
OY	2267	AGCTGACCAATTAGAAATACCATATGAGAGCAATATCTAGAGTA/-AATGAGAC	2346
Db	2098	AGCTGACCAAAATTAGAAATACCATATGAGAGCAATATCTAGAGTA/-AATGAGAC	2157
OY	2327	GAGGAATATCTTTACAAAAACAGCAGACAGTGCATATGGCTTTCCAG/AGATTTAG	2386
Db	2158	GAGGAATATCTTTACAAAAACAGCAGACAGTGCATATGGCTTTCCAG/AGATTTAG	2217
OY	2387	AAAAATGAAAAATTTGCTGAAGTAGAGATCCAAACCCACCATCTGATGGATCTGACACA	2446
Db	2218	AAAAATGAAAAATTTGCTGAAGTAGAGATCCAAACCCACCATCTGATGGATCTGACACA	2277
OY	2447	GCAATGATTCMAAACCCATGACACAGAAATGAAACAAAAGAAATCATTA/TAATTTGCA	2506
Db	2278	GCAATGATTCMAAACCCATGACACAGAAATGAAACAAAAGAAATCATTA/TAATTTGCA	2337
OY	2507	CTGGAATAATATCTGCTTATGCTTACCACTGCGACAAAGAGTCTCATATTAAAG	2566
Db	2338	CTGGAATAATATCTGCTTATGCTTACCACTGCGACAAAGAGTCTCATATTAAAG	2397
OY	2567	AATGTAACATGTTATCCGTTATGTCGTGTCACCAATGAATAGCAATCTCTCGAGCCGC	2626
Db	2398	AATGTAACATGTTATCCGTTATGTCGTGTCACCAATGAATAGCAATCTCTCGAGCCGC	2457
OY	2627	GTGGTCGAGCCAGAGCTGATGAGAGCACTACCTCTGTTGTCATCACTCTTGAGAG	2686
Db	2458	GTGGTCGAGCCAGAGCTGATGAGAGCACTACCTCTGTTGTCATCACTCTTGAGAG	2517
OY	2687	TTATGCAACATGAGACAGTTATATGATTTCCGACAGAAATGATATTAATCTATACAT	2746
Db	2518	TTATGCAACATGAGACAGTTATATGATTTCCGACAGAAATGATATTAATCTATACAT	2577
OY	2747	GTGTCAAAATATGAACCGAGAGATGTCCTATAGATTTTGGATTTAGATGCAAA	2806
Db	2578	GTGTCAAAATATGAACCGAGAGATGTCCTATAGATTTTGGATTTAGATGCAAA	2637
OY	2807	GATATATGAAAAAATAAATAAACCAGAGAAATATTCGAAGCATTAAGATTAAC	2866
Db	2638	GATATATGAAAAAATAAATAAACCAGAGAAATATTCGAAGCATTAAGATTAAC	2697
OY	2867	CATCACTAATACCTTCCTTGCAAAAACTGCAGCTATGCTACCTGTTCTGAGAGATA	2926

[illegible]

PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226686.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0233297.
 PR 14-SEP-2000; 2000US-0233398.
 PR 14-SEP-2000; 2000US-0233399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239335.
 PR 13-OCT-2000; 2000US-0239337.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.

PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246539.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0251989.
 PR 06-DEC-2000; 2000US-0251479.
 PR 06-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251866.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 (HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Barash SC, Ruben SM;
 WPI: 2001-465566/50.
 P-PSDB: AAU23090.
 Novel polypeptides and polynucleotides useful for diagnosing,
 preventing, treating neural, immune system, muscular, reproductive,
 pulmonary, cardiovascular, renal, proliferative disorders and cancerous
 diseases
 Claim 4: SEQ ID NO 186; 1180bp; English.
 The present invention relates to the isolation of novel human enzyme
 polypeptides (AAU2915-AAU23814), and the cDNA and genomic sequences
 encoding them. The enzyme polypeptides of the invention may comprise the
 functional classes of oxidoreductases, transferases, hydrolases, lyases,
 isomerases or ligases. The sequences of the invention are useful in the
 diagnosis, treatment, prevention and/or prognosis of a wide range of
 disorders including hyperproliferative disorders (e.g. cancer),
 immunodeficiency disorders (e.g. AIDS) autoimmune disorders
 (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),
 metabolic disorders (e.g. phenylketonuria), inflammatory disorders
 (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis).

CC blood-related disorders (e.g. haemophilia), reproductive disorders
 CC (e.g. infertility) and infectious disorders (e.g. Influenza). The
 CC polynucleotides of the invention can also be used in gene therapy.
 CC AAS40785-AAS41684 represent cDNA sequences encoding for the novel human
 CC enzyme polypeptides of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPRO
 CC at ftp.wipro.int/pub/published_pcl_sequences.
 XX

Sequence 1967 BP: 723 A: 344 C: 416 G: 476 T: 8 other:

Query Match 42.2%; Score 1419; DB 22; Length 1967;

Best Local Similarity 99.9%; Pred. No. 0; Mismatches 1; Indels 0; Gaps 0;

Matches 1469; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1047 AGAGAAATGTGCGACGAGAGATCCCGGAGCAGACACCTCCAGCTCAGCTTTACCAAT 1106
 DB 2 AAGAAATGTGCGACGAGAGATCCCGGAGCAGACACCTCCAGCTCAGCTTTACCAAT 61
 OY 1107 GGAAATGTGCGACGAGAGATCCCGGAGCAGACACCTCCAGCTCAGCTTTACCAAT 1166
 DB 62 GGAAATGTGCGACGAGAGATCCCGGAGCAGACACCTCCAGCTCAGCTTTACCAAT 121
 OY 1167 TGGAAATGTGCGACGAGAGATCCCGGAGCAGACACCTCCAGCTCAGCTTTACCAAT 1226
 DB 122 TGGAAATGTGCGACGAGAGATCCCGGAGCAGACACCTCCAGCTCAGCTTTACCAAT 181
 OY 1227 ATCTGAGCTGGAAAGTATAGTCTTCTCAATAGGACTGCTAGCTTACCAAT 1286
 DB 182 ATCTGAGCTGGAAAGTATAGTCTTCTCAATAGGACTGCTAGCTTACCAAT 241
 OY 1287 CCGCAAGAGTCCAGACCTTTTGAAGAAATGATGCTGATAGGAAAGTCTGA 1346
 DB 242 CCGCAAGAGTCCAGACCTTTTGAAGAAATGATGCTGATAGGAAAGTCTGA 301
 OY 1347 TACCAATGTAAATATATCTTCCAGAGTGTCAAGCTCTGATATATATCACTAG 1406
 DB 302 TACCAATGTAAATATATCTTCCAGAGTGTCAAGCTCTGATATATATCACTAG 361
 OY 1407 AGCTCAATCTTGAAGAACTCCCTCTTAACTTGAAGAAATGAGAGATCTGTGTCTCA 1466
 DB 362 AGCTCAATCTTGAAGAACTCCCTCTTAACTTGAAGAAATGAGAGATCTGTGTCTCA 421
 OY 1467 ATGTGAGCTTTCCTCATTTATCATGTGATGATGATGATGATGATGATGATGATGAT 1526
 DB 422 ATGTGAGCTTTCCTCATTTATCATGTGATGATGATGATGATGATGATGATGATGATGAT 481
 OY 1527 GTATATATCATCATGAGCATTTATTTGATGAGAGATGAAAGAAATGATGATGATGATGAT 1586
 DB 482 GTATATATCATCATGAGCATTTATTTGATGAGAGATGAAAGAAATGATGATGATGATGAT 541
 OY 1587 AGAAATCAAAACAGATATTCCTCTCCAGATAGTGGAGATGAGATGATGATGATGATGAT 1646
 DB 542 AGAAATCAAAACAGATATTCCTCTCCAGATAGTGGAGATGAGATGATGATGATGATGAT 601
 OY 1647 TGGAGGGGCGCAGAGCAAGCCAAAGCTGAAGACACATTTTAAACTATGTCCTCAATCT 1706
 DB 602 TGGAGGGGCGCAGAGCAAGCCAAAGCTGAAGACACATTTTAAACTATGTCCTCAATCT 661
 OY 1707 TGAATGATTTACTATTAATACTGTAAAGAAACCTTGAATCAATGAAATGATGATGATGAT 1766
 DB 662 TGAATGATTTACTATTAATACTGTAAAGAAACCTTGAATCAATGAAATGATGATGATGAT 721
 OY 1767 GGAAGCATGCAAGAGTTGGCATTCAGATGACACAGAGAGATGATGATGATGATGATGAT 1826
 DB 722 GGAAGCATGCAAGAGTTGGCATTCAGATGACACAGAGAGATGATGATGATGATGATGAT 781
 OY 1827 ACTTGAAGAAATATGACAAAGATTCAACTTATGTCAAATGATGATGATGATGATGATGAT 1886
 DB 782 ACTTGAAGAAATATGACAAAGATTCAACTTATGTCAAATGATGATGATGATGATGATGAT 841
 OY 1887 TGGAACTCAACCTATGACAAATGAGGCAATTCAAATGAGAAAGAAAGCTTCAAAAAAGG 1946
 DB 1887 TGGAACTCAACCTATGACAAATGAGGCAATTCAAATGAGAAAGAAAGCTTCAAAAAAGG 1946

DB 842 TGGAACTCAACCTATGACAAATGAGGCAATTCAAATGAGAAAGAAAGCTTCAAAAAAGG 901
 OY 1947 AAATGCAAAAGAAAGCTTTTGGAGACATTTGAGAGATGAGAGATGAGAGATGAGAGATGAGAGAT 2006
 DB 902 AAATGCAAAAGAAAGCTTTTGGAGACATTTGAGAGATGAGAGATGAGAGATGAGAGATGAGAGAT 961
 OY 2007 TAATGACAAATTCGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2066
 DB 962 TAATGACAAATTCGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1021
 OY 2067 GAAAGATAGAGATTTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2126
 DB 1022 GAAAGATAGAGATTTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1081
 OY 2127 TTGTATGAT 2186
 DB 1082 TTGTATGAT 1141
 OY 2187 TAGATTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2246
 DB 1142 TAGATTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1201
 OY 2247 CCGAGATATGAAATGAAAGCTGACCAATTAAGAAATGATGATGATGATGATGATGATGATGATGATGAT 2306
 DB 1202 CCGAGATATGAAATGAAAGCTGACCAATTAAGAAATGATGATGATGATGATGATGATGATGATGATGAT 1261
 OY 2307 TAGGACTGAGGATGAGACGAGAGATTAATCTTTCAAAAAACGACAGAGATGATGATGATGATGAT 2366
 DB 1262 TAGGACTGAGGATGAGACGAGAGATTAATCTTTCAAAAAACGACAGAGATGATGATGATGATGATGAT 1321
 OY 2367 GCTTCCCGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2426
 DB 1322 GCTTCCCGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1381
 OY 2427 TCTGATGAGCTGAGACAGCAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2486
 DB 1382 TCTGATGAGCTGAGACAGCAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1441
 OY 2487 AGTCATTAATTAATTTGAGCTGAGAGAAAT 2516
 DB 1442 AGTCATTAATTAATTTGAGCTGAGAGAAAT 1471

RESULT 5
 ABA04916
 ID ABA04916 standard, DNA: 1443 BP.
 XX
 AC ABA04916;
 XX
 DT 05-MAR-2002 (first entry)
 XX
 DE Human RNA helicase Rhl16 related DNA sequence #3.
 XX
 KW Human: Rhl16; RNA helicase; cytosolic; virucide; anti-HIV;
 KW immunosuppressive; immunostimulatory; antihemagic; antidiabetic;
 KW antidiabetic; osteoporosis; antidiabetic; antidiabetic; antidiabetic;
 KW antidiabetic; cancer; infection; HIV; hepatitis; genetic disease;
 KW autoimmune disease; graft rejection; vaccine; ds.
 XX
 OS Homo sapiens.
 OS
 PN W0200185955-A1.
 PN
 PD 15-NOV-2001.
 PD
 PF 11-MAY-2001; 2001WO-FR01441.
 PF
 PR 11-MAY-2000; 2000FR-0006030.
 PR
 XX (ISTA-) ISTAC.
 XX (INSP-) INST PASTEUR LILLE.
 PA
 PI Bahr G, Cocude C, Capron A;

PD 15-NOV-2001.
 XX
 PE 11-MAY-2001; 2001WO-FR01441.
 XX
 PR 11-MAY-2000; 2000FR-0006030.
 XX
 PA (ISTA-) INSTAC.
 XX (INSP-) INST PASTEUR LILLE.
 PI Bahr G, Cocude C, Capron A;
 XX
 DR WPI: 2002-082898/11.
 XX
 PT New polypeptide, useful for treating and diagnosing cancer or
 PT inflammation, and drug screening, comprises a human polynucleotide
 PT homologous to RNA helicase.
 PS
 XX Example 2: Page 93-94; 114pp; French.
 CC The present invention relates to human RH16 (see AAM47798). RH16 is a
 CC 116kDa protein and has homology to RNA helicases (DEXH box). RH16 and
 CC its coding sequence are useful for treating cancer; acute or chronic
 CC infections (especially by HIV or hepatitis B or C); inherited genetic
 CC diseases; (auto)immune diseases (particularly rheumatoid arthritis,
 CC arteriosclerosis, osteoporosis and diabetes; but many others listed) and
 CC to prevent graft rejection. RH16 and its coding sequence are also useful
 CC for inducing, or increasing, the immune response to a vaccine. The
 CC present sequence was used in an example from the present invention.
 SO
 Sequence 1284 BP; 489 A; 198 C; 261 G; 335 T; 1 other:
 Query Match 32.9%; Score 1106; DB 24; Length 1284;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 156; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

DB 541 TGATGAGACACCTACGTCCTGTTGCTACAGAGTGTTCAGAGTATTCAGACGTGAGAC 600
 QY 2703 AGTTATGATTTCCGAGAGAGAGATGATGATTAAGATGATGATTTGTTCAAAATAGAA 2762
 DB 601 AGTTATGATTTCCGAGAGAGAGATGATGATTAAGATGATGATTTGTTCAAAATAGAA 660
 QY 2763 ACCAGAGATGATGATGATGATTTGGAATGATGATGATGATGATGATGATGATGATGAT 2822
 DB 661 ACCAGAGATGATGATGATGATTTGGAATGATGATGATGATGATGATGATGATGATGAT 720
 QY 2823 AATGAAAACCAAGAAATATGTCAGACATTAAGAAATTAAGAAATTAAGAAATTAAGAAAT 2882
 DB 721 AATGAAAACCAAGAAATATGTCAGACATTAAGAAATTAAGAAATTAAGAAATTAAGAAAT 780
 QY 2883 CCTTTGCAAAAACCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2942
 DB 781 CCTTTGCAAAAACCTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
 QY 2943 AATGATCAGCTCAATATGACCCAGAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAAT 3002
 DB 841 AATGATCAGCTCAATATGACCCAGAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAAT 900
 QY 3003 AGCACTGCAAAAAGAGTGTGCGACATTAAGAAATTAAGAAATTAAGAAATTAAGAAATG 3062
 DB 901 AAGCACTGCAAAAAGAGTGTGCGACATTAAGAAATTAAGAAATTAAGAAATTAAGAAATG 960
 QY 3063 CCAGCTTGGGGAACATGATGCTGCAAAAGCTTGAATTTGCTGCTGCTGCTGCTGCTGCTG 3122
 DB 961 CCAGCTTGGGGAACATGATGCTGCAAAAGCTTGAATTTGCTGCTGCTGCTGCTGCTGCTG 1020
 QY 3123 GAATTTTACTGCTTTCAAAAATTAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATG 3182
 DB 1021 GAATTTTACTGCTTTCAAAAATTAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATG 1080
 QY 3183 ATTACTATCAATTTCCCAATCTTGAATTAAGAAATTAAGAAATTAAGAAATTAAGAAAT 3242
 DB 1081 ATTACTATCAATTTCCCAATCTTGAATTAAGAAATTAAGAAATTAAGAAATTAAGAAAT 1140
 QY 3243 TTAGACTGTAAGAAATTTTAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATG 3302
 DB 1141 TTAGACTGTAAGAAATTTTAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATG 1200
 QY 3303 ATTAAATGATTCATTAATCTACAGAACTGACATTAAGAAATTAAGAAATTAAGAAATG 3361
 DB 1201 ATTAAATGATTCATTAATCTACAGAACTGACATTAAGAAATTAAGAAATTAAGAAATG 1259

RESULT 7
 ID AAI59285
 AA159285 standard; cDNA: 1557 BP.
 XX
 AC AAI59285;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polynucleotide SEQ ID NO 1488.
 XX
 KW Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemoretic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia; ss.
 XX
 OS Homo sapiens.
 XX
 FN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 XX

Novel polypeptides and polynucleotides useful for diagnosing

DR P-PSDB: AAM41915.

XX Novel nucleic acids and polypeptides, useful for treating disorders

PT such as central nervous system injuries -

XX Claim 1: SEQ ID NO 5060; 10078bp; English.

PS The invention relates to human nucleic acids (AA157798-AA161369) and
XX the encoded polypeptides (AAM38642-AAM42213) with neurotropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC activation of the activities such as: immune system suppressor,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammatory leukemias and
CC C.N.S disorders.
CC Note: the sequence data for this patent did not form part of the printed
CC specification.

XX Sequence 1382 BP; 513 A; 217 C; 285 G; 367 T; 0 other:

Query Match 28.6%; Score 961; DB 22; Length 1382;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 1061; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2252 AATATGAAATGAAAGAGTACCAATTAAGAAATACATAATGAGCAAT ATACTAGCA 2311
DB 265 AATATGAAATGAAAGAGTACCAATTAAGAAATACATAATGAGCAAT ATACTAGCA 324
QY 2312 CTGAGCAATCAGCAGAGCAATATCTTACAAAACAGCAGAGAGTGC ATGCGCTTT 2371
DB 325 CTGAGCAATCAGCAGAGCAATATCTTACAAAACAGCAGAGAGTGC ATGCGCTTT 384
QY 2372 CCCAGTGAATTAAGAAATGAAAGTGTGAGTACAGTCAAGTCCCATCTGA 2431
DB 385 CCCAGTGAATTAAGAAATGAAAGTGTGAGTACAGTCAAGTCCCATCTGA 444
QY 2432 TTGAGCTGAGCAGCAGCAGTGTCAAAACCATGACAGAAATGAAATGAGCAAT 2491
DB 445 TTGAGCTGAGCAGCAGCAGTGTCAAAACCATGACAGAAATGAAATGAGCAAT 504
QY 2492 TTAGTAATTTGCGACTGAGAAATCATCTGCTATGCTACACAGTTCAGAGAG 2551
DB 505 TTAGTAATTTGCGACTGAGAAATCATCTGCTATGCTACACAGTTCAGAGAG 564
QY 2552 GTCGTGATTAAGAAATGAAATGAAATGAAATGAAATGAAATGAAATG 2611
DB 565 GTCGTGATTAAGAAATGAAATGAAATGAAATGAAATGAAATGAAATG 624
QY 2612 CCATGCTCCAGGCCCGCTGTCGAGCCAGCAGTGTAGAGCACTACGTTGGTGTG 2671
DB 625 CCATGCTCCAGGCCCGCTGTCGAGCCAGCAGTGTAGAGCACTACGTTGGTGTG 684
QY 2672 ACAGTGTTCAGAGAGTTATGCAACATGAGACAGTTAATGATTTCAGAGTATGATCT 2731
DB 685 ACAGTGTTCAGAGAGTTATGCAACATGAGACAGTTAATGATTTCAGAGTATGATCT 744
QY 2732 ATTAAGCTTACATGTTGTTCAAAATATGAAACAGAGAGTAATCTCATATGATTTTGG 2791
DB 745 ATTAAGCTTACATGTTGTTCAAAATATGAAACAGAGAGTAATCTCATATGATTTTGG 804
QY 2792 AATTACAGATGCAAGATATATGAGAAAGAAATGAAACCAATGAAATATGCAATC 2851
DB 805 AATTACAGATGCAAGATATATGAGAAAGAAATGAAACCAATGAAATATGCAATC 864
QY 2852 ATTACAGATATACCATCATGTAATTAATCTTCTTGGAAAAATGCAATGCTGCTAGCT 2911
DB 865 ATTACAGATATACCATCATGTAATTAATCTTCTTGGAAAAATGCAATGCTGCTAGCT 924

QY 2912 GTTCTGGGAGAGATATCCATGTAATTTAGAAAATGATCAGCTCATATATACCCAGAAAT 2971
DB 925 GTTCTGGGAGAGATATCCATGTAATTTAGAAAATGATCAGCTCATATATGACCCAGAAAT 984
QY 2972 TCAAGCACTTTACATGTTAGAGAAAACAAAGCAGCTGCAAAAGAGTGTGCGACTATC 3031
DB 985 TCAAGCACTTTACATGTTAGAGAAAACAAAGCAGCTGCAAAAGAGTGTGCGACTATC 1044
QY 3032 AAATTAATGTTGAATATCTCTGCAAAATGTGCGCAGCTTGGGGAACAATGATGTGCACA 3091
DB 1045 AAATTAATGTTGAATATCTCTGCAAAATGTGCGCAGCTTGGGGAACAATGATGTGCACA 1104
QY 3092 AAGGCTGATTTGCTGCTGCTCAAAATTAAGAAATTTGTGTGTTTCAAAATTAAT 3151
DB 1105 AAGGCTGATTTGCTGCTGCTCAAAATTAAGAAATTTGTGTGTTTCAAAATTAAT 1164
QY 3152 CACAAAGAAACATATACAAAAGTGGTAGAATTAACATACATATCCCAATCTGACT 3211
DB 1165 CACAAAGAAACATATACAAAAGTGGTAGAATTAACATACATATCCCAATCTGACT 1224
QY 3212 ATTGAAATGCTGTTTATTGATGATGAGATTAGCACTTATGAAATCTTTTAAA 3271
DB 1225 ATTGAAATGCTGTTTATTGATGATGAGATTAGCACTTATGAAATCTTTTAAA 1284
QY 3272 TACTATCAGTTAAACATTTAATGATTAATGATTAATGATTAATGATTAATGATTAATG 3314
DB 1285 TACTATCAGTTAAACATTTAATGATTAATGATTAATGATTAATGATTAATGATTAATG 1327

RESULT 10
AAS01149
ID AAS01149 standard; DNA; 1258 BP.
XX
AC AAS01149;
XX
DF 12-SEP-2001 (first entry)
XX
DE Interferon induced nucleic acid, IFN4.
XX
XX Interferon induced nucleic acid; autoimmune disease; lupus erythematosus;
KM immunodeficiency; stroke; acquired immunodeficiency syndrome; AIDS;
KM graft rejection; viral infection; hepatitis; aplastic anaemia; cancer;
KM human immunodeficiency virus; HIV; immune-mediated glomerulonephritis;
KM haematologic disease; chronic neutropenia; myocardial infarction;
KM neurological disease; Alzheimer's disease; Parkinson's disease; tumour;
KM amyotrophic lateral sclerosis; spinal muscular atrophy; human; IFN4; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
ET CDS 1..915
ET /tag= a
ET /product= "IFN4"
FT /note= "Interferon induced polypeptide"
PN WO200118208-A2.
XX
PD 15-MAR-2001.
XX
PE 08-SEP-2000; 2000WO-US24704.
XX
PR 08-SEP-1999; 99US-0152921.
PR 20-OCT-1999; 99US-0160575.
PR 20-JAN-2000; 2000US-0177104.
PR 07-SEP-2000; 2000US-0656633.
XX
PA (CURA-) CURAGEN CORP.
XX (BIOJ) BIOGEN INC.
XX
PI Peyman JA, Da Silva A, Hochman P, Hsu A;
XX WPI; 2001-235201/24.

PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -

XX Example 4: SEQ ID NO: 9283; 650bp + Sequence Listing: English.

XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.

XX Sequence 595 BP; 141 A; 162 C; 154 G; 134 T; 0 other;

Query Match 15.0%; Score 505; DB 22; Length 595;
Best Local Similarity 100.0%; Pred. No. 6.1e-233;
Matches 505; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 ACAACAGCAGCATCTGCTTGGGAGAAACCTCTCCCTCTCTGAGAAAGATATGTCGAA 176
DB 595 ACAACAGCAGCATCTGCTTGGGAGAAACCTCTCCCTCTCTGAGAAAGATATGTCGAA 536
QY 177 TGGGATTTCCACAGAGAGAAATTTCCGCTATCTCATCTCTGCTTCAAGGCTGAA 236
DB 535 TGGGATTTCCACAGAGAGAAATTTCCGCTATCTCATCTCTGCTTCAAGGCTGAA 476
QY 237 AATGTACATCCAGGTGGAGAGCTGTGCTGACTACCTGACCTTTCTGCTTACAGAGTGA 296
DB 475 AATGTACATCCAGGTGGAGAGCTGTGCTGACTACCTGACCTTTCTGCTTACAGAGTGA 416
QY 297 GGAGCAGATTGAGAGAGAGAGTGGCAGCTCCGGGAACATGAGGAGATTAAGTCTGCTG 356
DB 415 GGAGCAGATTGAGAGAGAGAGTGGCAGCTCCGGGAACATGAGGAGATTAAGTCTGCTG 356
QY 357 GAGCAGCTTGGAGAGAGAGAGTGGCAGCTTGGAGAGAGTGGCAGCTTGGAGAGAG 416
DB 355 GAGCAGCTTGGAGAGAGAGAGTGGCAGCTTGGAGAGAGTGGCAGCTTGGAGAGAG 296
QY 417 CCGGAGAAACCGGAGAGAGTGGCAGCTTGGAGAGAGTGGCAGCTTGGAGAGAG 476
DB 295 CCGGAGAAACCGGAGAGAGTGGCAGCTTGGAGAGAGTGGCAGCTTGGAGAGAG 236
QY 477 CTTCTCATGTTTGGAGAGAGTGGCAGCTTGGAGAGAGTGGCAGCTTGGAGAGAG 536
DB 235 CTTCTCATGTTTGGAGAGAGTGGCAGCTTGGAGAGAGTGGCAGCTTGGAGAGAG 176
QY 537 CACTCTGTGGAGAGAGCTTCTAGTTAGAGAGCTTGTGATAGTGCATGTAAGAGAACT 596
DB 175 CACTCTGTGGAGAGAGCTTCTAGTTAGAGAGCTTGTGATAGTGCATGTAAGAGAACT 116
QY 597 GTTGACAATTGAAGACAGAAACCGG 621
DB 115 GTTGACAATTGAAGACAGAAACCGG 91

RESULT 15

AAK35181/c
ID AAK35181 standard; DNA; 595 BP.

XX AAK35181;

XX 06-NOV-2001 (first entry)

XX Human bone marrow expressed single exon probe SEQ ID N 9738.

XX Human bone marrow expressed exon; gene expression analysis; probe;

XX microarray; cancer; leukemia; lymphoma; myeloma; ss.

XX Homo sapiens.

XX WO200157276-A2.

PD 09-AUG-2001.

XX 30-JAN-2001; 2001MO-US00668.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488900/53.

XX Human genome-derived single exon nucleic acid probes useful for

XX analyzing gene expression in human bone marrow -

XX Example 4: SEQ ID NO: 9738; 658bp + Sequence Listing: English.

XX The present invention provides a number of single exon nucleic acid

XX probes which are derived from genomic sequences expressed in the human

XX bone marrow. They can be used to measure gene expression in bone marrow

XX samples, which may enable the improved diagnosis and treatment of cancers

XX such as lymphoma, leukemia and myeloma. The present sequence is one of

XX the probes of the invention.

XX Sequence 595 BP; 141 A; 162 C; 154 G; 138 T; 0 other;

Query Match 15.0%; Score 505; DB 22; Length 595;

Best Local Similarity 100.0%; Pred. No. 6.1e-233;

Matches 505; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 ACAACAGCAGCATCTGCTTGGGAGAAACCTCTCCCTCTCTGAGAAAGATATGTCGAA 176
DB 595 ACAACAGCAGCATCTGCTTGGGAGAAACCTCTCCCTCTCTGAGAAAGATATGTCGAA 536
QY 177 TGGGATTTCCACAGAGAGAAATTTCCGCTATCTCATCTCTGCTTCAAGGCTGAA 236
DB 535 TGGGATTTCCACAGAGAGAAATTTCCGCTATCTCATCTCTGCTTCAAGGCTGAA 476
QY 237 AATGTACATCCAGGTGGAGAGCTGTGCTGACTACCTGACCTTTCTGCTTACAGAGTGA 296
DB 475 AATGTACATCCAGGTGGAGAGCTGTGCTGACTACCTGACCTTTCTGCTTACAGAGTGA 416
QY 297 GGAGCAGATTGAGAGAGAGTGGCAGCTCCGGGAACATGAGGAGATTAAGTCTGCTG 356
DB 415 GGAGCAGATTGAGAGAGAGTGGCAGCTCCGGGAACATGAGGAGATTAAGTCTGCTG 356
QY 477 CTTCTCATGTTTGGAGAGAGTGGCAGCTTGGAGAGAGTGGCAGCTTGGAGAGAG 536
DB 235 CTTCTCATGTTTGGAGAGAGTGGCAGCTTGGAGAGAGTGGCAGCTTGGAGAGAG 176
QY 537 CACTCTGTGGAGAGAGCTTCTAGTTAGAGAGCTTGTGATAGTGCATGTAAGAGAACT 596
DB 175 CACTCTGTGGAGAGAGCTTCTAGTTAGAGAGCTTGTGATAGTGCATGTAAGAGAACT 116
QY 597 GTTGACAATTGAAGACAGAAACCGG 621
DB 115 GTTGACAATTGAAGACAGAAACCGG 91

Sat May 17 10:42:58 2003

us-09-515-363c-1.rng

Search completed: May 17, 2003, 03:39:19
Job time : 775 secs

• • •